Early Growth Technology Analysis: Case Studies in Solar Energy and Geothermal Energy

by

Ayse Kaya Firat

M.Eng.
Texas A&M University, 2006

B.S.
Bogazici University, 2004

Submitted to the Engineering Systems Division
in Partial Fulfillment of the Requirements for the Degree of
Master of Science in Technology and Policy

at the
Massachusetts Institute of Technology

September 2010

Signature of author: ____________________
Technology and Policy Program, Engineering Systems Division
August 27, 2010

Certified by: ____________________
Stuart Madnick
John Norris Maguire Professor of Information Technology, MIT Sloan School of
Management & Professor of Engineering Systems, MIT School of Engineering
Thesis Supervisor

Accepted by: ____________________
Dava J. Newman
Professor of Aeronautics and Astronautics and Engineering Systems
Director, Technology and Policy Program

[back of title page]

Early Growth Technology Analysis: Case Studies in Solar Energy and Geothermal Energy

by

Ayse Kaya Firat

Submitted to the Engineering Systems Division
on August 27, 2010 in Partial Fulfillment of the Requirements For the Degree of
Master of Science in Technology and Policy
at the Massachusetts Institute of Technology

ABSTRACT

Public and private organizations try to forecast the future of technological developments and allocate funds accordingly. Based on our interviews with experts from MIT's Entrepreneurship Center, Sloan School of Management, and IBM, and review of literature, we found out that this important fund allocation process is dominated by reliance on expert opinions, which has important drawbacks alongside its advantages.

In this Thesis, we introduce a data-driven approach, called early growth technology analysis, to technology forecasting that utilizes diverse information sources to analyze the evolution of promising new technologies. Our approach is based on bibliometric analysis, consisting of three key steps: extraction of related keywords from online publication databases, determining the occurrence frequencies of these keywords, and identifying those exhibiting rapid growth. Our proposal goes beyond the theoretical level, and is embodied in software that collects the required inputs from the user through a visual interface, extracts data from web sites on the fly, performs an analysis on the collected data, and displays the results. Compared to earlier software within our group, the new interface offers a much improved user experience in performing the analysis.

Although these methods are applicable to any domain of study, this Thesis presents results from case studies on the fields of solar and geothermal energy. We identified emerging technologies in these specific fields to test the viability of our results. We believe that data-driven approaches, such as the one proposed in this Thesis, will increasingly be used by policy makers to complement, verify, and validate expert opinions in mapping practical goals into basic/applied research areas and coming up with technology investment decisions.

Thesis Supervisor: Stuart Madnick

John Norris Maguire Professor of Information Technology, MIT Sloan School of Management & Professor of Engineering Systems, MIT School of Engineering

ACKNOWLEDGMENTS

I am grateful to my Thesis advisor and a true leader, Prof. Stuart Madnick, for his guidance, support, and patience as I traveled in the thorny path of scholarship. Prof. Madnick taught me how to ask the right questions, and showed immense understanding and patience during my difficult times. Thanks to Prof. Woon of Masdar Institute for Science and Technology (MIST) from whom I have received salient advice and feedback; and to Skype for making our virtual meetings possible.

The fine staff, faculty, and students of the Technology and Policy Program kept me focused and joyful during my time at MIT. Special thanks to Sydney Miller, and Ed Ballo of TPP. Thanks to my fellow students Blaine Ziegler, and Steven Camina with whom I have been able to talk about my ideas and receive helpful feedback whenever I hit a dead end. I am also thankful to Allen Moulton, Michelle Cole, and Michael Siegel from the COIN Group.

I am grateful to Prof. Howard Anderson, Prof. Bill Aulet, and Satwik Seshasai, who agreed to interview with me, and occupied a Chapter in my Thesis.

Finally, I would like to acknowledge my inspirational parents Necla and Halil Kaya, amazing siblings Hatice and Mikail Kaya, and my compassionate husband, Aykut Firat, who encouraged me along the entire path, never showed anything but confidence in me, and taught me the most important things I know.

TABLE OF CONTENTS

TABLE OF FIGURES

LIST OF TABLES

CHAPTER 1: INTRODUCTION

1.1 MOTIVATION

Identifying novel technologies that have the potential to generate high commercial returns at an early stage is critical to venture capital firms, entrepreneurs, and policy makers alike. Decision makers and strategy setters have to track current state of research by sifting through volumes of data, and weigh future scenarios by seeking in-depth expert opinions. The Internet is both a friend and foe in this process, as it offers ever-growing amounts of information for richer yet more time-consuming analysis. Clearly, new automated methods are needed to aid decision makers in this challenging process.

1.2 EARLY GROWTH TECHNOLOGY ANALYSIS

In this study, we investigate the efficacy of one such automated method in identifying potentially promising set of technologies in a given area of interest using online databases of scientific publications. We provide a set of software tools that implement this method, and we present two case studies by applying the methodology to the fields of solar and geothermal energy. We evaluate the utility of this automated approach by conducting interviews with subject-area experts and noting their reactions.

We call the method used in this Thesis, Early Growth Technology Analysis (EGTA), since we try to locate technologies that are in the early, low-prevalence phase of their life cycle and are likely to exhibit high growth. With EGTA, we take advantage of the knowledge buried in online scientific publications to perform bibliometric analysis, consisting of three key steps:

1. Term collection by extracting related keywords from articles for a given area of interest
2. Determining the occurrence frequencies of these keywords (hit counts)
3. Identifying those exhibiting rapid growth, particularly if starting from a low base

Term collection starts with a seed term such as "solar energy" that acts as a proxy for the general technology area of interest. We then utilize online publication databases such as Compendex, Inspec, and Scirus to find terms relevant to this seed term. Some of these online databases are freely available (e.g. Scirus), some require subscription (e.g. Compendex and Inspec), and yet others require permission for programmatic access (e.g. Google Scholar) (See Appendix 1 for a detailed description of these sources).

After collecting the related terms from a set of sources, the next step is to find out hit counts of each term over a number of years. For this task, we use the hit counts returned by a set of data sources, which in our case are the same sources we use for collecting terms. Finally, we narrow down our original list to a subset of terms that seem to fit the early growth description by ranking our list using these hit counts.

1.3 PURPOSE: IDENTIFY CANDIDATE RESEARCH TOPICS TO CONSIDER

The top technologies produced by this method are not necessarily "the best" of their kind or "the answer" to a particular question, but merely are interesting emerging directions for decision makers to consider. To be more concrete, our algorithm is not going to help organizations like Masdar or IBM, sponsors of our research, to decide which specific projects to fund, which is Phase 3 of the overall funding process as shown towards the right in Figure 1. Both Masdar and

IBM have their own decision criteria to work on this step such as quality of research, past track record of the researchers/employees, cost-benefit analysis, expert opinions, and fit with organizational goals. Our goal is to help the decision makers in the former step, Phase 2 (see Figure 1, Left), that involves going from a broad area of interest such as energy or photovoltaics to a few possible themes that are believed hold a lot of promise. With the help of EGTA, decision makers will only face a modest list of promising ideas, or in other terms, options to consider. The decision makers may already have their own 'promising ideas' list, and may use our results to add to that list or validate their options. After this initial step, these decision makers may use criteria of their choice to make their final decision.

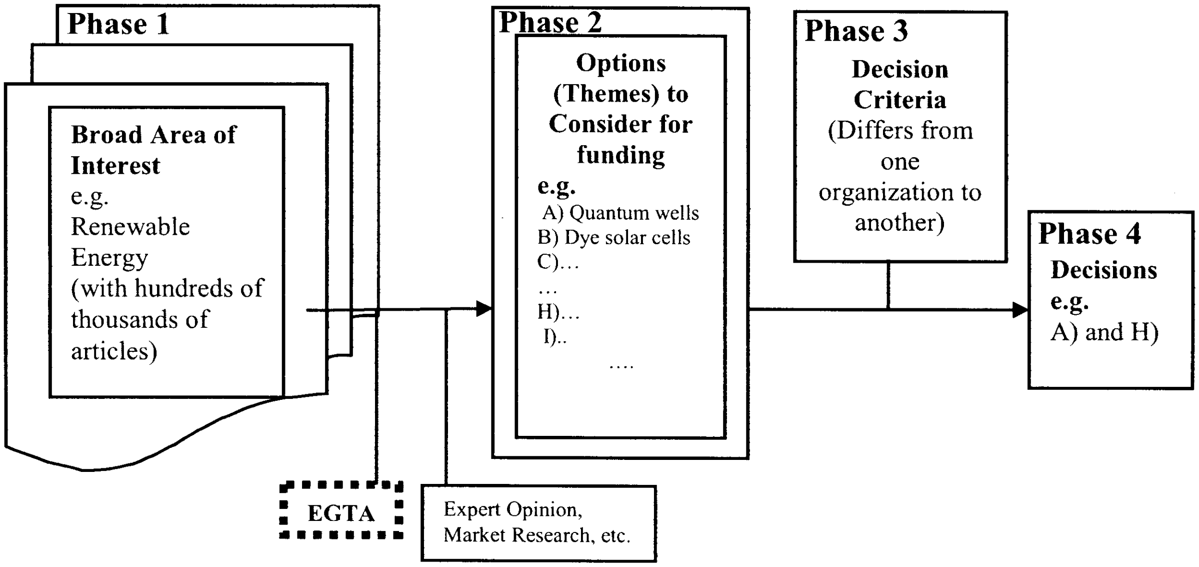


Figure 1: The process decision makers go through for investment/funding allocation decisions

1.4 RELATIONSHIP TO TRADITIONAL METHODS

Our interviews with entrepreneurs and venture capitalists, and the review of literature on technology investments reveal that decision makers widely rely on the subject area experts' predictions in both of the decision steps discussed in Figure 1. The first of these two steps (narrowing down a broad area of interest to a few promising options) requires an in-depth knowledge of the area as well as its relationship with other fields. Although experts in a field could have a detailed understanding of their field, it would be unreasonable to expect them to have exhaustive knowledge of every facet of their field. Emerging technologies, in particular, are unfortunately the ones of which they are most likely unaware. With this project, our main contribution is to help decision makers and experts identify emerging technologies in a field they are interested in; ensure that they do not miss an important development; augment and validate their already established ideas on options to consider for investment and fund allocation.

Apart from this contribution, we also advance the state of data collection within our group with a new set of software tools. Although effective in producing results, earlier software developed in our group by fellow researchers were not initially designed with "user friendliness" or ease of extendibility in mind. The new set of tools, including the Hit Aggregator, Cameleon Scheduler and Early Growth Technology Analysis (EGTA) tool, all have visual interfaces.

Moreover, extending the analysis with new publication databases can be done without any coding. The most significant software, EGTA tool, was built using the Model-View-Controller (MVC) design pattern, which makes further development much easier.

1.5 CASE STUDIES IN SOLAR AND GEOTHERMAL ENERGY

To provide a focus for subsequent discussions and evaluations, we conducted two case studies in solar and geothermal energy by using our set of software tools. These two case studies and our interviews with subject area experts indicate that our proposed approach can be an important decision aid for policy makers.

We believe that data-driven approaches, such as the one proposed in this Thesis, will increasingly be used by policy makers to complement, verify, and validate expert opinions in mapping practical goals into basic/applied research areas and coming up with funding allocation and technology investment decisions.

1.6 THESIS STRUCTURE

This chapter has presented a brief overview and the objective of our project, along with a brief description of the methods and tools used to achieve these objectives. Chapter 2 explains the policy motivation for the project followed by a Literature Review in Chapter 3, which presents a review of the research and literature in the fields of technology forecasting, and bibliometrics.

Chapter 4, Early Growth Technology Analysis, describes how we extract keywords from online publication search engines, how the associated publication counts are extracted, and how we use these counts to identify fast-growing technologies. It also provides a comparison of the data gathering approaches adopted within our research group, and why we believe the approach presented in this Thesis is superior.

Chapter 5, Tool Development, describes the software tools we created to make our approach more "user friendly" in detail.

Chapter 6 and 7, Solar Energy and Geothermal Energy Case Study Results, presents results of our software to analyze the fields of solar and geothermal energies.

In Chapter 8, we discuss our interviews with subject-area experts and how they view the method we outlined in this Thesis, before concluding in Chapter 9, with a summary of the main findings of the project and areas left for future research.

CHAPTER 2: POLICY MOTIVATION

Technology policy makers widely rely on subject area experts in making their funding allocation decisions. Experts, with their experience and domain knowledge, are invaluable resources in helping decision makers, but even they may not be fully aware of all the promising developments in broad and complex fields of technology. In the fast paced Internet age, it is no longer possible for a human to wake up every morning and read every blog post, news and journal article in their domain of interest and not miss an important development before it becomes a headline. There is an obvious need for a computer to process vast amounts of relevant information out there, and present a summary in human digestible format.

One of our primary goals in this thesis is to create and experiment with an automated data-driven method that helps experts acquire a more complete understanding of their area of expertise. The promise of this goal can be better understood, if the reader is first told how these decisions are currently made in policy circles. We start with an anecdote, and continue with a background on the policy making process in the rest of this chapter.

2.1 KURZWEIL AND ISRAEL'S ENERGY POLICY

I am at "The VC Portfolio in 2030", a panel discussion organized by the MIT Sloan Venture Capital Club. World-renowned futurist, author, and inventor, Ray Kurzweil, is discussing his controversial views of the immediate future and its dramatic implications to our everyday life and taking questions from two technology investors to determine the best way to deploy capital in a future that we can't yet imagine.

Kurzweil is an avid proponent of regularity in technological progress, a hard core techno-determinist. According to Kurzweil, the biggest problem today is that people do not recognize this regularity in technological progress enough. But if they did, if they had a sense and vision to recognize the power of technologies, this would eliminate impediments to further development, adoption and diffusion. Not recognizing these changes prevents harvesting the benefits rapidly and adequately.

In today's discussion, Kurzweil talks about investing in renewable energy. He says the field of energy is being transformed by exponential growth, just like information technology (IT) has been for several decades. In IT, most of us are familiar with the Moore's law that describes a long-term trend in the history of computing hardware, in which the number of transistors that can be placed inexpensively on an integrated circuit has doubled approximately every two years. Energy is currently a field dominated by fossil fuels, which is a 19th century technology. Kurzweil's claim on doubling the performance of clean energy technologies every two years, as the semiconductor industry has seen with Moore's Law, seems like a tough goal to hit.

Solar To Rule in the Next 16 Years

Kurzweil mentions a recent study where a panel of experts including Google Co-Founder Larry Page and Kurzweil himself convened by the National Academy of Engineering. The panel investigated all the emerging energy technologies and picked solar as having the most potential because of its applicability to nanotechnology. The reason why solar energy technologies will advance exponentially, Kurzweil says, is because it is an "information technology" (one for which we can measure the information content), and thereby subject to the "Law of Accelerating

Returns". Solar and wind power currently supply about one percent of the world's energy needs, but advances in technology are about to expand with the introduction of nano-engineered materials for solar panels, making them far more efficient, lighter and easier to install.

"We also see an exponential progression in the use of solar energy," he says. "It is doubling now every two years. Doubling every two years means multiplying by 1000 in 20 years. At that rate we'll meet 100 percent of our energy needs in the next 16 years."

A Conversation Changing a Country's Energy Technology Roadmap

Right after these statements, comes the most interesting part of the Kurzweil speech. Kurzweil says he shared his views on solar energy with the Prime Minister of Israel a couple of months ago at the Israeli Presidential Conference. He reports that the Prime Minister said "Well, that's great. Is there enough sun light to double eight more times?" Kurzweil explained to him that there is 10,000 times more sunlight than we need to meet 100 percent of our energy needs: "We could place the efficient solar panels 16 years from now on just a couple percent of the world's unused lands and meet all of the world's energy needs"

Based on their conversation, Kurzweil says, the Prime Minister announced a day later a 10-year-plan harvesting the best scientists and engineers in Israel to completely replace fossil fuels with solar energy not just for Israel but for the world. Kurzweil said he told the Prime Minister afterwards that this was overly optimistic as it is going to take at least 16 years.

As a student of Technology Policy, I am quite surprised by the power and consequences of a single conversation. Even Kurzweil himself confesses that he was surprised. This anecdote is a vivid example of the enormous role expert opinions play in guiding decision makers in public and private organizations.

Next, we explore why policy makers rely so much on unstructured processes by understanding the difficulty of setting research and investment priorities.

2.2 DIFFICULTY OF SETTING RESEARCH AND INVESTMENT PRIORITIES

Policy makers at public and private organizations try to forecast the future of technological developments and are instructed to allocate funds to different areas of research using "practical objectives or goals serving as a motivation" (Government Performance and Results Act, GPRA of 1993). In research and development, however, "practical objectives or goals" can be very ambiguous. How exactly do these decision makers and organizations, then, go about mapping practical goals into applied research topics?

According to Ken Oye, MIT Professor of Political Science and Engineering Systems, such mapping decisions drive policy makers insane more than anything else. Oye mentions the struggle of Larry McCray, currently at MIT's Program On Emerging Technologies (POET), with such mapping decisions when he was heading the Policy Division of the National Research Council (NRC). As part of his job, McCray was asked to instruct and guide federal activity on shaping research priorities at the NRC. He had to provide objective guidance on the processes through which American research priorities should be set. Larry McCray was bothered because he concluded he did not have much in the way of good guidance to offer to the federal government on how to set research priorities with reference to practical objectives.

The National Science and Technology Council (NSTC) guidance or Government Performance and Results Act (GPRA) of 1993 are all these bits and pieces of regulation that are results oriented. Decision makers look at a program or a broad field and try to see what results would follow from it to give more money to promising programs or emerging technologies in a broad field.

Companies face a similar challenge like governments. How are corporate decisions, such as investing millions of dollars to a new technology in the field solar energy, being made? Are funding allocation decisions based on "objective, repeatable, and quantifiable" decision parameters? As Jerome C. Glenn, the director of the Millennium Project indicates in his reply to the above-mentioned questions, "Corporations tend not to share that information." Alan Porter, a leading figure in technology forecasting (TF), states that "These decisions are most often based on 'tacit knowledge' without much systematic TF or competitive technical intelligence (CTI) being utilized."

Let us briefly look into some public and private organizations such as European Commission, National Science Foundation, IBM and Novartis to understand how they set their research and investment priorities.

Public Institutions

European Commission

The European Commission determines research priorities by sending a questionnaire to a panel of about 1300 experts in all countries of the enlarged Europe. These experts represent the totality of European countries and the science and technology fields. Two thirds of the experts interviewed are from public research while the remaining one third comes from the private sector. Most of the participants are high level experts and more than two thirds of the experts interviewed are directors / heads of department in their organization [European Commission 2006].

European commission faces several difficulties in this process. The directors in the participating organizations might have prior interests and be inclined to advancing the technologies they are interested in. Timelines projected by directors may be very different from what the people working underneath the director are thinking. Furthermore, the composition of experts because of country quota reasons may diminish the quality of the panel.

National Science Foundation (NSF)

NSF apportions its funds by creating panels of reviewers. What follows is a description of going from Phase 3 to Phase 4 mentioned earlier in Figure 1. We believe, based on informal talks, NSF uses a similar procedure in going from Phase 1 to Phase 2 of the funding process.

First, fund seekers complete proposals and send them back to NSF. NSF creates a panel from a list of reviewer applicants to assess the proposals. To understand the potential difficulties of this process, we need to look at the composition of expert reviewers in these panels. Are people that seek out the reviewing positions necessarily the people that are most oriented towards the latest advances? The ones who volunteer might have prior interests, and be inclined to interpret knowledge to make projections and evaluate risks to benefit themselves. Consequently, claims of interpretative authority may influence fund allocation decisions possibly in a non-optimal way when there is no data-driven mechanism to substantiate expert claims [National Science Foundation 2010].

Private Institutions

Novartis

Novartis' approach to setting research priorities and investments at the corporate level is a good example of the strong trend in systematic technology intelligence (TI) undertaken in many technology-intensive large companies [Lichtenthaler 2004a], [Lichtenthaler 2004b]. Novartis uses 180 globally distributed participants, including specialist teams, informal discussion groups and several fulltime technology intelligence specialists, to communicate during the year via intranet, where new trends are discussed. Furthermore, three to four times a year they meet physically in order to integrate the information gathered into a holistic and shared picture and in order to create an atmosphere of trust with their colleagues. This process somewhat resembles IBM's Horizon Watch.

IBM

HorizonWatch[1] is an internal IBM 'Grass-Roots' community that has been in place for over nine years. HorizonWatch Community was started to provide executive, strategy, and marketing teams with an early warning identification system of new, emerging opportunities, threats and trends in the marketplace. The community has over 1800 members from all types of functions, all divisions, and all geographies in IBM. Within IBM, and HorizonWatch, there are also employees whose full time job is to research, analyze, and write about emerging topics that will have an impact on IBM's ability to grow.

HorizonWatch community members are interested in learning, and collaborating on emerging business issues, trends and technologies. They meet via conference calls. Topics are presented to the community by subject matter experts. In between conference calls, they collaborate via the HorizonWatch blog, which is open to IBM employees only. The community has evolved into a collaborative network of people who are interested in hearing about and discussing emerging technology topics.

GlaxoSmithKline

GlaxoSmithKline faced the research priority and investment setting problem immediately after the merger of SmithKline & French with Beecham (1990) [Norling et al. 2000]. The new corporate senior management faced the challenge of reallocating the combined $1 billion R&D budget.

After the merger, a team from R&D and central marketing was assembled to look at the company's existing portfolio of therapeutic area research. The goal was to assess the viability of each research area and to explore new areas of unmet medical conditions or needs that could be profitably explored. A consulting firm, which had also been called upon, proposed locating the various therapeutic areas within a typical positional map of commercial attractiveness vs. technical feasibility or strength (see Figure 2, left). The firm was unwilling to recommend the discontinuation of work in any one of the therapeutic areas.

The head of R&D, however, was not satisfied, and asked the small, four-person intelligence group in the R&D section to look at another tool to guide the company in refocusing its R&D resources. That led to the application of scientometrics[2] or science mapping, a technique of using computer algorithms to identify connection patterns within the recently published scientific literature.Based on these patterns, a structural map of the scientific community can be created, showing the interrelationships between disciplines and the distribution of research communities.

[1] http://horizonwatching.typepad.com/horizonwatching/2007/04/the_horizonwatc.html

[2] Scientometrics is concerned with the quantitative features and characteristics of science. Emphasis is placed on investigations in which the development and mechanism of science are studied by statistical mathematical methods. In practice, as in the SmithKline Beecham case, scientometrics is often done by measurement of (scientific) publications using bibliometrics. For purposes of this report, scientometrics and bibliometrics are used interchangeably.

A scientometric or knowledge map can identify the structure of a particular area of scientific research and measure its performance: How "hot" is this research area? How quickly are new discoveries being made? Is the field growing, or imploding upon itself? Maps can be drawn for each level in the hierarchy and color-coded according to performance measures.

SmithKline Beecham used this technique as one element in the redirection of its R&D resources. After generating scientometric maps of the seven research-based universes (or therapeutic areas) in which the merged company was active, they concluded that the field of gastrointestinal disease research in particular was not generating a significant amount of high-performance research. The positional map was redrawn (Figure 2, right). The company decided to close its research activities in this area, and to focus on research in the remaining six: the central nervous system, inflammatory disorders, cardiorespiratory disorders, metabolic disease, cardiovascular disease and anti-infection agents. The company then turned its attention to research platform (technology) areas, identifying networks of research communities common to the seven therapeutic areas. One such network constituted a technology universe working in the broad area of genomics, an interesting but uncertain field in the early 1990s.

Through scientometrics, it identified several university groups and small companies that were conducting high-momentum research in the genomics area. Further investigation of these high-momentum groups led to the first genomics agreement in the industry between SmithKline Beecham and Human Genome Sciences. Scientometrics also helped SmithKline Beecham to locate a multimillion-dollar research facility focusing on the central nervous system. Maps showed that centers of excellence in CNS research were located on the east and west coasts of the U.S. and in France, which was where the company ultimately built one of its research satellites. In short, scientometric technology gave the company an important intelligence perspective that enabled it to reshape its research portfolio for greater productivity, and to define a number of promising technology opportunities.

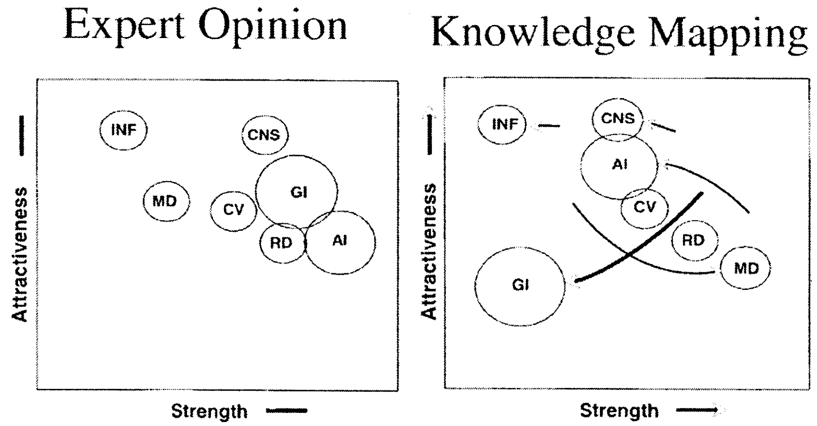


Figure 2: The map (left) of commercial attractiveness vs. technical strength for seven therapeutic areas — the central nervous system (CNS), inflammatory disorders (INF), cardiorespiratory disorders (RD), metabolic disease (MD), cardiovascular disease (CV), gastrointestinal disease (GI), and anti-infection agents (AI)—was redrawn based on the use of scientometrics (right). GI was then dropped from the R&D program [Norling et al. 2000].

Other technology intensive companies

In a recent study by [Lichtenthaler 2004b], a total of 147 interviews were performed, in 26 technology intensive large companies in Europe and North America (Table 1). Interviewed were specialists of the technology intelligence units and the technology acquisition intelligence units, as well as customers of these intelligence units from top management including in each case: the head of research or the chief technology officer, a member of middle management and a few individual researchers. Companies from the pharmaceutical, telecommunication equipment and automotive/machinery industries were examined with the goal of exploring industry differences in the management of technology intelligence processes.

	Pharmaceuticals	Telecommunications equipment	Automobile/Machinery	Total
Europe	Novartis	Nokia	Sulzer	17
	Roche	Ascom	DaimlerChrysler	
	Bayer	Siemens	Hilti	
	Zeneca	Swisscom	Schindler	
	Boehringer Ingelheim	Philipps	Landis & Gyr	
	Hoechst Marion Roussel		Bosch	
USA	Pfizer	Lucent Technologies	Ford	9
	Merck	Nortel Networks		
	Glaxo Wellcome[1]	Cisco		
	SmithKline Beecham			
	DuPont			
Total	11	8	7	26

Table 1: The companies included in "Technology Intelligence Processes in Leading European and North American Multinationals." study [Lichtenthaler 2004b]

According to this study, the selection of the TF methods in a company were influenced by the objective of individual or organizational learning sought, time horizon of planning, and industry. Table 2 shows the intensity of different information source use in the industries studied.

	Pharmaceuticals	Electronics	Auto/Machinery
Publication frequency analyses	• • •	• •	•
Publication citation analyses	• • •	–	–
Quantitative conference analyses	• •	• • •	•
Patent frequency analyses	• •	• • •	• • •
Patent citation analyses	–	–	•
S-curve analyses	–	–	–
Benchmarking studies	• • •	• • •	• • •
Portfolios	• • •	• • •	• • •
Delphi studies	–	–	–
Expert panels	• • •	•	• •
Flexible expert interviews	• • •	• • •	• • •
Technology roadmaps	• •	• • •	–
Product technology roadmaps	–	• • •	•
Product roadmaps	• • •	–	–
Experience curves	•	• • •	• •
Simulations	• •	–	–
Option pricing models	• •	–	–
Scenario analyses	• • •	• • •	• • •
Lead user analyses	–	• • •	• •
Quality function deployment	–	• •	• • •

• • • = often used   • • = sometimes used   • = rarely used   – = not used

Table 2: Intensity of use of different information sources in the industries studied [Lichtenthaler 2004b].

Take the science-driven pharmaceutical industry as an example. Starting from fixed customer needs, which can be determined in the form of long-term epidemiological studies, the scientific environment is scanned for most promising innovations. New scientific research results are often of high competitive relevance and are immediately used. Publication citation analyses are therefore quite important in the pharmaceutical industry.

Many pharmaceutical companies combine publication citation analyses with an iterative marketing process of ideal product identification.. As the projects move forward in the product pipeline, techno-economic, time and competitive aspects especially start to dominate assessments. Quantitative assessments are increasingly used. Pharmaceutical companies try to handle the technological uncertainty and the high failure rate of R&D projects by using options pricing methods. The large R&D budgets and the rising pressure to increase effectiveness in the selection of R&D projects are the root cause of the use of expensive and complex methods, such as simulations and publication citation analyses.

On the other hand, the telecommunications equipment industry is a market-driven industry. Technological progress and market development are closely coupled. This is reflected by the importance of lead user analyses, technology product roadmaps and scenario analyses. The integrated technology and market planning is seen as necessary because of the high rate of technological and market change. In the automotive industry in contrast, there is a slow rate of technological and market change.

In the telecommunications equipment industry, normally several technologies compete to become a standard and often imply different markets. At the same time, these technologies are only unstable dominant designs, which are substituted after a comparably short time. Besides the identification of innovation impulses from science, monitoring of the changing techno-economic importance in order to select the right technology and the right time to invest in a technology is of great importance. The importance of the monitoring of the techno-economic changes is mirrored by the intensive use of quantitative monitoring of conferences, experience curves and patent frequency analyses. Publication citation analyses and patent citation analyses are not used because scientific advances often take many years to become competitively relevant and the rate of change is too fast.

The automobile and machinery industries are more mature and less dynamic industries than the pharmaceutical and telecommunications equipment industries. Technological as well as market uncertainty are comparably low. The main focus is on the integration of customer needs in products and incremental innovations. Radical innovations are mainly triggered by the regulatory environment. Very often, therefore, scenario analyses, quality function deployment and lead user analyses are used. Changes in the scientific environment are perceived to be of less competitive importance compared to the telecommunications equipment and pharmaceutical industries. Patent citation analyses are mainly used to scan for new technologies.

2.3 ANALYSIS AND CRITIQUE OF THE CURRENT APPROACHES IN USE

Although there are advanced data driven approaches undertaken especially in the private industry to determine research priorities and investment decisions, our interviews with subject-matter experts reveal that entrepreneurs and investment makers often employ similar approaches to organizations like European Commission, IBM and Novartis. When making investment decisions, they tend to rely on a small number of data sources, experts and friends. As discussed before, an expert may not be fully aware of all the promising developments in broad and complex fields of technology. It is hard to imagine an expert who can remain up to date in all relevant areas, process all the information out there and ensure not to miss an important development.

Furthermore, such decisions are prone to confirmation bias. Confirmation bias (also called confirmatory bias) is a tendency for people to favor information that confirms their preconceptions or hypotheses, independent of their truth. This results in people selectively collecting new evidence, interpreting evidence in a biased way, or selectively recalling information from memory. Instead of investigating in a neutral, scientific way, people tend to test hypotheses in a one-sided way, focusing on one possibility and neglecting alternatives. Wishful thinking and information processing limitations also contribute to the overall issue. Confirmation bias and overconfidence in personal beliefs strengthen beliefs in the face of contrary evidence, and can lead to disastrous decisions, especially in organizational, military and political contexts.

In policy making, we even have critiques who question the good will of experts and thus, the validity of expert opinions. Techno-constructivism is a school of thought composed of techno-constructivists exemplified by Paul Rabinow of University of California, Berkeley, an extremely distinguished anthropologist who has studied with the famous post-modernist critic, Michel Foucault. Rabinow sees these debates on technological forecasting, and efforts on prioritization of research and cost benefit analyzes as fundamentally corrupt. He argues that the claims of expertise are often amplified, exaggerated, and manipulated to claim power in debates [Rabinow 2004].

Availability of data is also an important factor in technological forecasting. Levary and Han [Levary and Han 1995] argues that given a small amount of low or medium validity data, and no similarity between proposed technology and existing technologies, a reasonable choice is a method based on information obtained from a panel of experts (i.e., Delphi method or interviews). Given a moderate to large amount of medium to high validity data and a high degree of similarity between proposed technology and existing technologies, they propose using correlation analysis. When there is medium or large amount of high validity data, trend analysis is the most appropriate method. Yet, in practice, independent of these factors, expert opinion affects funding allocation decisions more than any other method.

Time horizon is also a critical factor in how organizations go about setting research priorities and funding allocation. According to [Lichtenthaler 2004b], the longer the time horizons, the more the studied companies try not to forecast the development of a technology as precisely as possible, but rather tend to determine a commonly shared and supported, partly normative future, starting from an intensive analysis of the environment.

Organizational attempts to base funding allocation decisions on "objective, repeatable, and quantifiable" decision parameters are usually unsuccessful. In a 2001 study, [Reger 2001] made interviews in 25 multinational companies[3]. More than half the firms investigated emphasized that

[3] Fifteen of the companies interviewed were in the fields of computers, electronics, energy, or aviation, and four companies in the automobile industry. The telecommunication/network operators sector was represented by four companies and the chemical industry by three. Sixteen of the corporations in the survey have their headquarters in Western Europe, five in Japan and five in the United States. The following persons were interviewed within the companies:

- The head of technology foresight, or those responsible for technology foresight processes,
- Heads of the technology planning/technology strategy group or department,

technology intelligence is an unstructured and unsystematic process – which illustrates the opportunity for improvement.

Our goal with this thesis is not to reshape the process of setting research priorities and investment decisions, but to offer a method that complements an expert's ability to predict the future of technological developments. With the tools we developed, experts will acquire a more complete and up to date understanding of their area of expertise. Our data-driven tools will help decision makers and experts identify emerging technologies in a field they are interested in, and ensure that they do not miss an important development in its early stage.

• Customers' such as, e. g. the head of an R&D/technology centre or the head of corporate research, the head of technology development in a business field or a member of a strategic committee.

Most interviews were conducted with senior managers responsible for the technology foresight process or for corporate R&D/technology strategy. All companies interviewed described their competitive environment as highly dynamic. The budget for research and development (R&D) in the interviewed firms was between 80 million and 4. 5 billion Euro.

CHAPTER 3: LITERATURE REVIEW

Our research is part of the "Technology Forecasting using Data Mining and Semantics" (TFDMS) [Woon & Madnick 2008(2] project undertaken collaboratively by Massachusetts Institute of Technology (MIT) and Masdar Institute of Science and Technology (MIST). The methods we use in our research are derived from the field of technology forecasting (TF) and data mining, also known as tech mining. We, thus, first provide a review of the literature surrounding the field of technology forecasting in this chapter.

3.1 TECHNOLOGY FORECASTING

TF, in general, applies to all purposeful and systematic attempts to anticipate and understand the potential direction, rate, characteristics, and effects of technological change. It especially focuses on invention, innovation, adoption, and use of technology. One imperfect yet useful analogy for TF is weather forecasting: TF enables better plans and decisions. A good forecast can help maximize gain and minimize loss from future conditions. Additionally, TF is no more avoidable than weather forecasting. All people implicitly forecast the weather, for example, by choosing to wear a raincoat, or carry an umbrella. Any individual, organization, or nation that can be affected by technological change, inevitably engages in forecasting technology, explicitly or implicitly, with every decision that allocates resources to particular purposes.

Ability to forecast emerging technologies inform critical choices at organizations of all sizes, from large multinational unions, such as the European Union, to small start-up companies. Large organizations need TF to:

- Prioritize R&D,
- Plan new product development,
- Make strategic decisions on technology licensing, joint ventures, and so forth.

Small organizations depend on technological innovation for their existence. In these companies, TF methods are used to forecast adoption or diffusion of innovations, where parameters such as rate of imitation by other adopters or rate of response to advertising can be measured. TF studies in companies are often called Competitive Technological Intelligence (CTI or TI).

In addition to mapping out commercially viable roadmaps for technological development, the TF field includes more social and diffuse measurements as well. For example, governments use national foresight studies to assess the course and impact of technological change for the purposes of effecting public policy. This includes what is known as technology assessment (TA) or social impact analysis, which examines the likely long-term effects of technological development as its impact spreads throughout society.

Furthermore, technology foresight studies are used as an awareness-raising tool, alerting industrialists to opportunities emerging in science and technology, and alerting researchers to the social or commercial significance and potential of their work [Coates et al. 2001].

3.1.1 Forms of Technology Forecasting and Related Terminology

There are many overlapping forms of forecasting technological developments, such as technology intelligence, forecasting, roadmapping, assessment, and foresight. There has been little systematic attention to the conceptual development of the field as a whole. Since 2003, the Technology Futures Analysis Methods Working Group (TFAMWG) has sought to lay a framework from which to advance the processes and the methods. They combined different forms of technology forecasting studies under the term technology futures analysis (TFA) and classified different forms as follows [TFAMWG 2004]:

- **Gathering and interpreting information**: Technology monitoring, technology watch, technology alerts.
- **Converting that information into actionable intelligence**: Technical intelligence and competitive intelligence.
- **Anticipating the direction and pace of changes**: Technology forecasting.
- **Relating anticipated advances in technologies and products to generate plans**: Technology roadmapping.
- **Anticipating the unintended, indirect, and delayed effects of technological changes**: Technology assessment, and forms of impact assessment, including strategic environmental assessment.
- **Effecting development strategy, often involving participatory mechanisms**: Technology foresight, also national and regional foresight.

Many of these forms of forecasting use similar tools to accomplish similar ends. But there is a general tendency in government to use phrases that separate thought from action, such as "assessment" and "foresight," while in industry there is a tendency to use phrases that link thought and action, such as "roadmapping" and "competitive technological intelligence." There are cross-national differences as well, propelled by the differences of societal expectations from markets and governments. Industrial roadmapping, a largely private sector led initiative, originated and became prevalent in the United States, while foresight, a government sponsored activity, became the preferred alternative in Europe. These forms of forecasting—national technology foresight, roadmapping, and competitive technological intelligence—came into prominence at different times, and with relatively little effort to clarify their similarities and differences.

TF usually focuses on specific technologies, but sometimes the scope is more encompassing. A firm might roadmap a set of related technologies and products; an industry association might roadmap the gamut of emerging technologies potentially affecting its sector; or a nation could roadmap technologies across its economic base. For example, a U.S. semiconductor industry association roadmap, regularly updated to support industry planning, had as its early objective as regaining global market share in semiconductors. If semiconductor technologies were addressed in a national foresight study, the scope might also include the needs and capabilities of the relevant sciences at the input end, and the possible societal costs and benefits at the outcome end.

Methodologically, both national foresight studies and roadmapping usually bring together people representing different expertise and interests, and use instruments and procedures that allow participants to simultaneously adopt a micro view of their own disciplines and a systems view of overriding or shared objectives [TFAMWG 2004].

In this thesis, we use TF in its broadest sense covering all of the activities mentioned in the framework mentioned above.

3.1.2 Trend in TF Publications

How much TF research publication is out there? Figure 3 shows the results of querying Web of Science for "Technological forecasting" or "Technology forecasting". The activity seems encouraging for TF.

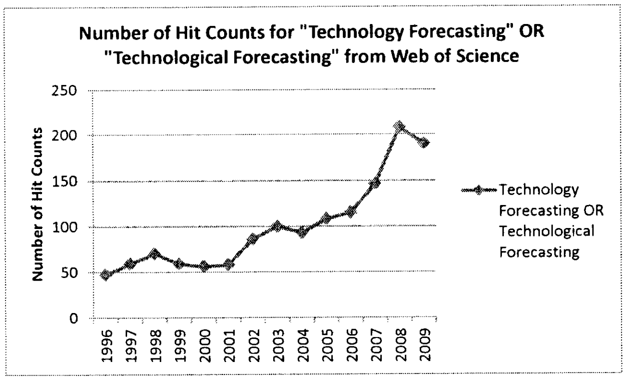


Figure 3: Number of hit counts for "Technology Forecasting" OR "Technological Forecasting from Web of Science between 1996 and 2009.

In 2006, Alan Porter prepared a literature profile of the TF domain helping to characterize the growing body of knowledge [Zhu and Porter 2002], [Porter 2007]. This study shows that the number of scholarly articles relating to TF is increasing. The study also examines the sectoral mix of institutions involved in TF as shown in Table 3 below. Note that the second grouping consolidates several difficult to distinguish types – governmental and non-governmental organizations, and other such institutes. Not surprisingly, publication of TF articles is strongly led by the academic community--which has the greatest stake in such publication-- but the substantial participation by government and industry is also notable.

Type	# of Articles	# of Authorships	% of Articles
Academic	567	779	58%
Gov't/NGO's/Institutes	174	210	18%
Industry	109	142	11%
Other	128	-	13%

Table 3: Leading Authoring Organizations by Sector [Porter 2007]

Where is TF work being published? Alan Porter's study lists 11 journals with 10 or more publications, where "Technological Forecasting & Social Change" is the leader, amidst strong representation from leading technology management journals (Table 4). The "Journal of Cleaner Production" focuses on sustainable development, while "Solid State Technology" shows a number of technology roadmapping articles.

Leading FTA Journals (# of Articles)
Technological Forecasting & Social Change (114)
International Journal of Technology Management (52)
Futures (49)
Research--Technology Management (26)
Abstracts of Papers, American Chemical Society (14)
Technovation (13)
Journal of Cleaner Production (12)
Journal of Forecasting (12)
R & D Management (11)
Solid State Technology (11)
Technology Analysis & Strategic Management (11)

Table 4: Leading FTA Journals [Porter 2007]

3.1.3 TF and Access to Information

Forecasters have long had complex algorithmic approaches at their disposal, but their ability to effectively execute those approaches has been limited by the availability of information and costs of manual information manipulation and analysis.

A defining characteristic of the Internet age has been the tremendously enhanced access to information. This offers particular promise to improve TF. There are many web sites that provide useful information, including projects, research opportunities, publications, citations, and patents.

Worldwide research and development activity results in explosive growth in the amount of scientific and engineering literature. For instance, Science Citation Index contains almost 15 million abstracts of journal and conference papers published since 1987. US, Japanese, and European patents are searchable online [Zhu & Porter 2002]. More importantly, many organizations license diverse R&D databases for unlimited searching, e.g., universities for their students and faculty.

3.1.4 TF Methods

There are hundreds of TF Methods, which can be fit into 9 families [Coates et al. 2001], [Gordon and Glenn 2003] as follows (areas marked with * show where our research fits in):

1) **Expert Opinion**
    - Delphi [iterative survey]
    - Focus Groups [panels, workshops]
    - Interviews
    - Participatory Techniques
2) **Trend Analysis**
    - Trend Extrapolation [Growth Curve Fitting]*
    - Trend Impact Analysis
    - Precursor Analysis
    - Long Wave Analysis
3) **Monitoring and Intelligence Methods**
    - Monitoring [environmental scanning, technology watch]
    - Bibliometrics [research profiling; patent analysis, text mining]*
4) **Statistical Methods**
    - Correlation Analysis
    - Demographics
    - Cross Impact Analysis
    - Risk Analysis
    - Bibliometrics [research profiling; patent analysis, text mining]*
5) **Modeling and Simulation**
    - Agent Modeling
    - Cross Impact Analysis
    - Sustainability Analysis [life cycle analysis]
    - Causal Models
    - Diffusion Modeling
    - Complex Adaptive System Modeling (CAS) [Chaos]
    - Systems Simulation [System Dynamics, KSIM]

- Technological Substitution
- Scenario-simulation [gaming; interactive scenarios]
- Economic base modeling [input-output analysis]
- Technology Assessment

6) **Scenarios**
- Scenarios [scenarios with consistency checks; scenario management]
- Scenario-simulation [gaming; interactive scenarios]
- Field Anomaly Relaxation Method [FAR]

7) **Valuing/Decision/Economics Methods**
- Relevance Trees [futures wheel]
- Action [options] Analysis
- Cost-benefit analysis
- Decision analysis [utility analyses]
- Economic base modeling [input-output analysis]

8) **Descriptive and Matrices Methods**
- Analogies
- Backcasting
- Checklist for Impact Identification
- Innovation System Modeling
- Institutional Analysis
- Mitigation Analysis
- Morphological Analysis
- Roadmapping [product-technology roadmapping]
- Social Impact Assessment
- Multiple perspectives assessment
- Organizational analysis
- Requirements Analysis [needs analysis]

9) **Creativity**
- Brainstorming [brainwriting; nominal group process (NGP)]
- Creativity Workshops [future workshops]
- TRIZ
- Vision Generation
- Science Fiction Analysis

We will briefly review some of the most popular methods in each category. Note that some of the methods fit into more than one family. For example, bibliometrics – which is a major focus of the MIT/MIST project – is listed under Trend Analysis, Statistical, and Monitoring/Intelligence Methods.

Expert Opinion

Expert Opinion methods include forecasting or understanding technological development via intensive consultation with subject-matter experts. The most popular method in this family is the Delphi Method. This method combines expert opinions concerning the likelihood of realizing the proposed technology as well as expert opinions concerning the expected development time into a single position. In Delphi, a sequence of individual interrogations is followed by information and opinion feedback derived from analyzing the initial response data. This feedback, which includes the reasoning and/or justification behind each individual expert's forecast, allows the other experts to revise their forecast in light of the new information. A single acceptable forecast is typically agreed upon after several rounds of this process [Levary and Han 1995]. Delphi, being the most widely used technique, has been subjected to scrutiny by many authors. Woundenberg, for instance, [Woundenberg 1991] questioned the accuracy and reliability of the Delphi method drawing upon the work of many other researchers like Campbell [Campbell 1966], Pfeiffer [Pfeiffer 1968], Dalkey [Dalkey 1969], Dalky and Helmer [Dalky et al. 1963], Farquhar [Farquhar 1970], Gustafson [Gustafson et al. 1973], Parente [Parente et al. 1984], Hill and Fowles [Hill and Fowles 1975] and Martino [Martino 1970].

Trend Analysis

Trend analysis involves prediction via the continuation of quantitative historical data into the future. Trend analysis is a broad term that encompasses economic forecasting models and techniques such as regression, exponential smoothing and Box-Jenkins' ARINA model and growth curve fitting [Levary and Han 1995]. A technology usually has a life cycle composed of several distinct stages. The stages typically include an initial adoption stage, a growth stage, a maturity stage and a declining final stage. Growth curve forecasting is based on the parameter estimation of a technology's life cycle curve. The growth curve forecasting method is helpful in estimating the upper limit of the level of technology growth or decline at each stage of the life cycle. This method of forecasting is also helpful in predicting when the technology will reach a particular life cycle stage.

One type of growth curve forecasting method is the Fisher-Pry Analysis. It is a mathematical technique used to project the rate of market adoption of technically superior new technologies and, when appropriate, to project the loss of market share by old technologies [Sahlo and Cuhls 2003]. The technique is based on the fact that the adoption of such new technologies normally follows the “logistic curve” pattern (also known as the S-curve), defined by two parameters. One of these parameters determines the time at which adoption begins, and the other determines the rate at which adoption will occur. These parameters can be determined from early adoption data, and the resulting pattern can be used to project the time at which market takeover will reach any given level. Results produced by this technique are highly quantitative. The technique is used to make forecasts such as how the installed base of telecommunications equipment will change over time, how rapidly a new chemical production process will be adopted, and the rate at which digital measuring devices will replace analog devices in petroleum refineries.

Monitoring and Intelligence Methods

Monitoring and its variations such as the Environmental Scanning and Technology Watch, are suitable for making one aware of changes on the horizon that could impact the penetration or acceptance of the technologies in the marketplace [Phillips et al. 2005]. Environmental scanning is considered as a central input to futures research [Woon & Madnick 2008(2)], but its output is seen as too general to support a specific decision. Its objectives then is to find early indications of important future developments to gain as much lead time as possible [Woon & Madnick 2008(2)].

Resource availability is one of the scoping issues associated with these methods since a number of the scanning approaches require the use of experts. Expert panels are created to look out for changes on the horizon that could be important to implement or accomplish plans. Experts are also tracked in a "scan the scanners" manner. TF analysts identify the experts in a field and keep track of those individuals by making occasional contact with them, observing them at conferences or searching the Internet for insights they may have posted.

Bibliometrics

Bibliometrics is a set of tools for analyzing publication data. Some of the bibliometric information associated with a publication includes author, affiliation, citations from other publications, co-citations with other publications, reader usage, and associated keywords. Bibliometrics can be used as a measure to describe research output of organizations, evaluate the impact of a technology or track the level of activity in a research field over time.

According to Porter and Cunningham [Porter and Cunningham 2005], "social scientists have applied methods of content analysis for decades. Counting scientific publication activity dates back at least to the pioneering work of Derek de Solla Price (1963)... With the advent of electronic text sources and analytical software, content analysis has matured into text mining... Data mining seeks to extract useful information from any form of data, but common usage emphasizes numeric data analysis...Text data mining or text mining exploits text sources of various sorts".

One of the most important aspects of bibliometric analysis is that it goes beyond the experts' biases, allowing the discovery of new facts and patterns that sometimes are not perceived due to the limit of knowledge or prejudiced visions. Some authors point out certain limitations of bibliometric analysis [Porter and Detampel 1995],[Porter 1998], considering that not all R&D activities are published or patented: much of the activity of technological development is not included either in journals, conferences, papers or patents in a timely fashion; the counting of publications does not distinguish the quality from its content; each institution has its own patenting policy; and there is no perfect system of classification and indexation of publications.

Besides these limitations, there are essential points for obtaining good results in text mining [de Miranda Santo et al. 2006]:

- **Knowledge of the subject under study**: it is important to have a good knowledge of the subject to define the search strategy in databases and analyze its results.
- **Knowledge of the databases to be used**: to know their contents and their structure, their level of standardization and the existing possibilities of data recovery are factors that define the success or failure of the task. The lack of standardization, for example, sometimes makes good text mining very hard due to low trustworthiness of the data.
- **Knowledge of patent information**: if patents are under study, it is important to know about the patents information structure, since they have rules of their own. Patent is a wide field, where techniques, products, applications and legal considerations are strongly mixed. This is also a field most of the time dedicated to industry people and, for example, the academic community does not cite patents very much. Nevertheless, patents are a unique source of information since most of the data and information published in patents are not published elsewhere.
- **Definition of search strategy**: it is an essential step and it is linked to the three previous ones, that is: knowledge of the subject and knowledge of databases and patents.

The use of restricted or extremely ample terms, for example, can lead to results that induce to errors of evaluation.

- **Usage of analytical tools**: it is important to have good text mining softwares and also to really know how to use them. Some commercial databases are beginning to provide analytical tools together with the search facilities, but they still have limited possibilities.
- **Results analysis**: experts must analyze the results trying to extract the best interpretation of the histograms, matrices and networks looking for strategic information.

The usage of text mining techniques must, necessarily, involve the experience of information professionals and of domain experts to be successful. The knowledge of information professionals on the available information sources, their contents and structure, and the opinion of experts to define the search strategy and to interpret the results are crucial for the quality of the final work.

Statistical Methods

In the Statistical Methods family, the most popular methods are correlation analysis and bibliometrics. Correlation analysis forecasts the development pattern of a new technology when the development pattern of the new technology is similar to those of existing technologies. Use of this method presupposes that data regarding the development patterns of the existing technologies are available [Phillips et al. 2005].

In 1983, when Martino published [Martino 1983], there was a correlation between the total installed steam turbine capacity in the United States and the maximum size of a single steam turbine electric generator. This would allow one to forecast the largest size of a steam turbine electric generator based on the forecast total industry capacity.

Many new or potential tools, currently used in future studies, have resulted from advances in information technology and science. Bibliometrics, and its specialized form scientometrics[4], for example, are two tools used traditionally by the information science experts to measure scientific productivity and to identify science and technology networks [de Miranda Santo et al. 2006].

Modeling and Simulation

A model is a simplified representation of the structure dynamics of some part of the real world. Models can exhibit future behavior of complex systems simply by isolating important system aspects from the inessential detail. Modeling requires a good understanding of interactions between these forecasts and the underlying variables or determinants.

One example in this family is agent modeling. An agent model involves the creation of computer generated "agents" that "populate" a computer screen, and interact with one another according to a set of behavioral rules [Gordon 2003]. The agents may be of different species; that is, they may have different attributes and may be assigned different rules. Their interaction over time is usually simulated by successive "plays" of the rules as the evolving attributes and spatial

[4] The term scientometrics is used to describe the study of science: growth, structure, interrelationships and productivity. [Hood and Wilson 2001] states that there has been considerable confusion in the terminology of the two closely related metric terms bibliometrics and scientometrics. Bibliometrics is a more general term referring to the statistical analysis of a document without the actual extraction of each document's fulltext. Scientometrics, on the other hand, is mainly used for the study of all aspects of the literature of science and technology, as the name would imply.

positions of the agents are computed. The spaces in the environment in which the agents are placed may also contain rules.

[Gordon 2003] describes an agent model that simulates the spread of an infection in a population, but explains that the model could be used to simulate any attribute that is passed from one person to others in society, such as a disease, an idea or belief, a fad, a market or a behavioral pattern. Using the model provided in the paper as a starting point, [Phillips et al. 2005] claims it may be possible to apply this concept to simulate the growth of use of sustainable energy technologies such as clean coal technology, tidal power and photovoltaics, if each sustainable energy technology is modeled as an agent. Each agent would have its own attributes and be governed by different rules. For example, an attribute of clean coal may be that it has a negative connotation while an attribute of photovoltaics may bet that it has a positive connotation while that of tidal power has a neutral connotation. Assuming that one set up the spaces to be in some way representative of society; filled with experts, private companies, "the public", government, etc., rules could be set up such that, for example, if a clean coal technology agent met a public space, the rule could require that the infection is retarded. However, if a clean coal technology agent met an expert space, the rule could require that the infection is advanced.

Systems simulation is another popular method in this family. The major benefit of systems simulation is to "allow users to search for the best approaches to an opportunity, facing a challenge, or solving a problem, without the risk or price of costly mistakes" [Woon and Madnick 2008(2)]. Given this benefit, it is possible to imagine configuring a system which contains all (or as many as reasonably possible) sustainable energy technologies and running a simulation to determine which technology will have the highest future value [Phillips et al. 2005]. However, although it is possible to imagine this model theoretically, the practical aspects of implementing such a model would be daunting because the accuracy of the pictures that system simulations create depends entirely on the quality of the data and on the realism of the way the relationships are expressed in the model. [Woon & Madnick 2008(2)].

Scenarios

Scenario writing proposes different conceptions of future technology. Each conception of the characteristics of the future technology is based on a well-defined set of assumptions. A scenario represents alternative characteristics of the future technology, with each alternative being based on certain assumptions and conditions. The forecaster evaluates the validity of the assumptions. The results of this evaluation are used to determine the scenario most likely to occur [Levary and Han 1995].

Most often, scenarios are used by top management to provide a better understanding of the range of possible business environments they must contend with in the future. As a tool for imagining alternative futures, scenario projects have helped many leaders gain perspective to guide their search for competitive advantage.

In the 1950s, Herman Kahn and his associates at the RAND Corporation adapted the meaning and method of theatrical scenarios to war planning [Millett 2003]. Kahn used scenarios to mean alternative paths resulting in alternative outcomes, such as his four scenarios of how nuclear war might erupt between the US and the Soviet Union.

Based on Kahn's work at RAND and later at his own Hudson Institute, war-planning scenarios were adapted by companies as a business planning tool in the early 1970s. Ian Wilson at GE, Pierre Wack at Shell, and Peter Schwarz at SRI International redefined scenarios as alternative outcomes of trends and events by a target year regardless of the precise sequence of events [Millett 2003]. Their scenarios were descriptions of future conditions rather than accounts of how events might unfold. Scenarios offered a set of distinct alternative futures to emphasize that the business environment was uncertain and could evolve in totally different ways. The scenarios provided a context for the development of long-term corporate strategic plans and near-term contingency plans. For example, Wilson led perhaps the first major corporate scenario project at GE that produced in 1971 four alternative scenarios of global and US economic and socio-political conditions in the year 1980: benchmark (with a 50 percent probability), more inward-looking societies (25 percent), more integrated societies (15 percent), and more disarrayed societies (10 percent) [Millett 2003].

Valuing/Decision/Economics Methods

The most popular method in this category is the "relevance tree approach". This is a normative approach to TF. The goals and objectives of a proposed technology are broken down into lower level goals and objectives in a tree-like format. In this way, the hierarchical structure of the technological development is identified. The probabilities of achieving the goals and objectives at the various levels of technological development must be estimated. The probabilities can then be used to forecast the likelihood of achieving the stated goals and objectives of the proposed technology [Levary and Han 1995].

Descriptive and Matrices Methods

A growing activity in this category is technology roadmapping, which projects major technological elements of product design and manufacturing together with strategies for reaching desirable milestones efficiently. Roadmaps typically run several technology or product generations (e.g., 2 to 10 years) ahead. In its broadest context, a science and technology roadmap provides a consensus view or vision of the future science and technology landscape available to decision makers. Thus, the predictive element emphasized in early TF is supplemented with a normative element, that is, narrower, more targeted, and more directly actionable than is the normative element implicit in TA. In the past, the institutional champions for roadmapping were mainly military industrial organizations; more recently, they have been other large corporations and industry associations [Coates et al. 2001].

UK Department of Trade and Industry 's Foresight Vehicle Technology Roadmap is a good example for technology roadmapping. The UK Foresight Vehicle roadmapping initiative involved 10 workshops over 10 months, with 130 people participating from over 60 organisations who tried to chart the future for road vehicles from a multistakeholder perspective in 2001. One of the technology elements of the roadmap shows how the fuel cell is expected to develop and challenge the internal combustion engine, highlighting how transitional phases involving hybrid vehicles may bridge the gap while the technology and necessary infrastructure develop. The overall roadmap provides a common framework and resource for the sector to collectively address the challenges facing the road transport system [Brown and Phaal 2001].

Analogies are also widely popular descriptive methods. The use of analogies in forecasting involves a systematic comparison of the technology to be forecast with some earlier technology that is believed to have been similar in all or most important respects. According to Martino in [Martino 1983], one of the shortcomings with analogies is that they "...are based on the assumption that there is a 'normal' way for people to behave and that given similar situations, they will act in similar ways. However, there is no guarantee that people today will act as people did in the model situation. Hence the forecast is at most probable, never certain".

3.1.5 How to Evaluate the Quality of TF Methods?

Evaluation of TF methods is quite challenging. Evaluation should establish, how much the method appears to be achieving its intended outcomes. Yet, there is no general-purpose toolkit for evaluating TF studies' influence and outcomes. A key challenge is establishing where a TF process begins and ends. Also, determining the extent to which an activity would have taken place without the intervention of the TF is problematic.

The Technology Futures Analysis Methods Working Group (TFAMWG) [TFAMWG 2004] gives a brief study that focuses on the evaluation of national Tech Foresight programs.[5] They choose Tech Foresight, as it has a mission of informing specific decisions. Tech Foresight also seeks to enlarge excessively short-term horizons and facilitate the formation of new networks around technologically and socially innovative activities.

TFAMWG [TFAMWG 2004] draws attention to two aspects of evaluation, product and process. *Product-oriented* work results, for example, in priority lists, reports arguing the case for a strategy in a particular field of science and technology (S&T), proposals for reform of educational systems, etc. It is possible to count and document products (reports, web pages, etc.), to examine their diffusion (readership, citations, etc.), and even to get some estimate of their use.

Process-oriented work results in network building, shared understanding, the formation of new alliances, bringing new participants into the innovation policy debate, etc. These consequences are harder to measure and monitor and will typically require more explicit examination - they will rarely be available as by-product data from the administration of a program.

TFAMWG [TFAMWG 2004] examines evaluation and use of Tech Foresight in terms of:

Futures: The assessment depends on the period that Tech Foresight addressed. In a short horizon (say, 5 years) critical technology exercise, this is not too serious a delay. But when Tech Foresight involves a time scale of 15 or more years, assessment is difficult—and its utility more problematic.

Participation and Networks: Examination of many aspects of the engagement of people in the Tech Foresight process and of the formation and consolidation of networks is best carried out in real time—memories get hazy rapidly and many of these activities go unrecorded. But many of the outputs and outcomes of such activities will take time to mature and require ex post investigation.

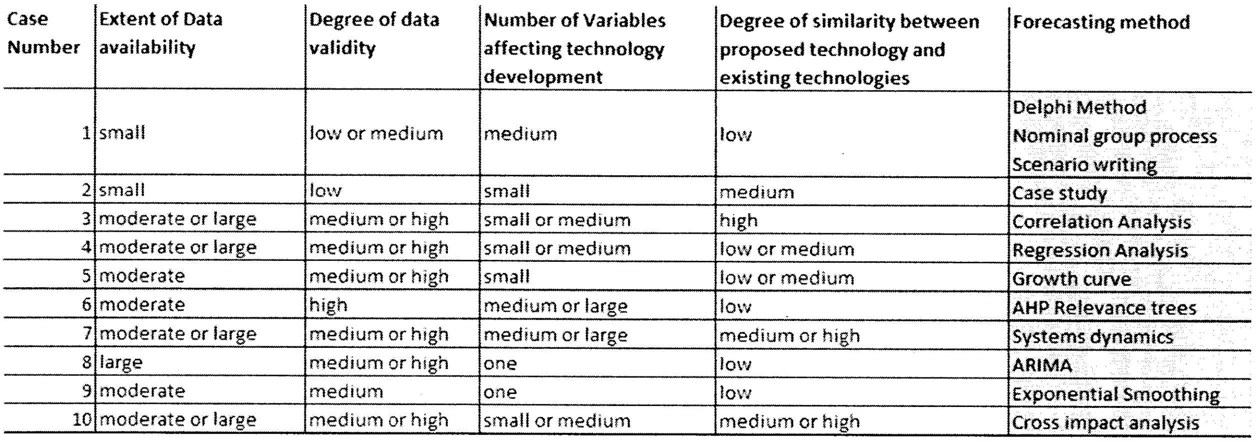
Action: A major question here is that of attribution. [TFAMWG 2004] claims that actions are often packaged as resulting from Tech Foresight, while in reality the decision makers use the reference to the study merely as a means of legitimation. Similarly, many actions may be taken that have their origins in the study but are not attributed to that source.

[5] Technology foresight is a term used for national TF activities in general. [Reger 2001] states that much of the pioneering work in technology foresight in the industry and at national level was done in the USA. Large think tanks, such as Rand and Hudson, made many technological forecasts since 1960s. The studies basically intended to help large corporations and government agencies to adjust their technological investment. Since the early 1970s, various ministries and agencies in Japan have been conducting repeated technological foresight studies (among them the Ministry of Trade and Industry (MITI) , Economic Planning Agency (EPA) and the Science and Technology Agency (STA)). Western European countries followed with systematic technology foresight activities in the 1990s [Farquhar 1970].

Choosing a Forecasting Method

A large number of methods have evolved for TF, but the quality of forecasts greatly depends on proper selection and application of appropriate methods. The application demands that the technique used need to be time-, space- and technology-specific. Yet, there is little research done on matching the TF methods techniques to a particular technology.

One such study comes from, Levary and Han [Levary and Han 1995], who have considered several basic factors such as the extent of data availability, the degree of data validity and degree of similarity between proposed technology and existing technologies. Each factor has been categorized into cases as small/low, medium/moderate, large/high and their combinations, and an appropriate forecasting method has been suggested. A summary of these suggestions are shown in Table 5 below.



Case Number	Extent of Data availability	Degree of data validity	Number of Variables affecting technology development	Degree of similarity between proposed technology and existing technologies	Forecasting method
1	small	low or medium	medium	low	Delphi Method Nominal group process Scenario writing
2	small	low	small	medium	Case study
3	moderate or large	medium or high	small or medium	high	Correlation Analysis
4	moderate or large	medium or high	small or medium	low or medium	Regression Analysis
5	moderate	medium or high	small	low or medium	Growth curve
6	moderate	high	medium or large	low	AHP Relevance trees
7	moderate or large	medium or high	medium or large	medium or high	Systems dynamics
8	large	medium or high	one	low	ARIMA
9	moderate	medium	one	low	Exponential Smoothing
10	moderate or large	medium or high	small or medium	medium or high	Cross impact analysis

Table 5: Forecasting methods for particular situations

A more recent study [Mishra et al. 2002] provides a comprehensive procedure to pick the right TF method. First they identify the characteristics of a technology that need to be considered (rate of change, ease of diffusion, number of alternatives available, etc). Next, using a 10-point scale, experts of the selected technology rate each of the characteristics for the selected technology. Then, using the same characteristics, experts of TF methods rate every method in the same manner. Finally, the profiles for the TF methods and technology profiles are superimposed to ascertain the ''best fit,'' i.e., the technique profile that closely matches the technology profile. By using this procedure, for example, they match normative technique to forecast Defense Weapon Systems, and Delphi method for IT (software for ecommerce).

An important element of the MIT/MIST TFDMS project is bibliometric analysis. As we defined earlier, bibliometrics is the statistical analysis of text documents, typically publications and patents. Since publications in this case refer mainly to academic publications and patents, science and technology intensive industries would logically be a better fit for this type of analysis. As patents and publications often deal with ideas and techniques in the relatively early stages of development, this is the stage at which bibliometric methods are most useful. Also, in the early stages of development, technical merit is probably the key determinant of success. Later on many other factors would influence the success of a technology or product, so there is a lot more complexity and noise. In such situations, "higher-level" features and pattern recognition techniques become more appropriate.

Many articles state that, because of the complexity of TF and because each forecasting method can deal with only limited aspects of a forecasting case, it is often advantageous to use several different forecasting methods simultaneously. In line with this, the MIT/MIST TFDMS research project extends and improves "tech-mining" techniques and introduces semantic enabled features. The performance of programs and tools are tested and fine-tuned with case studies on renewable energy and sustainability [Ziegler 2009], solar energy and geothermal energy.

3.1.6 Conclusion

This section presented many overlapping forms of forecasting technology developments and their impacts, including technology intelligence, forecasting, roadmapping, assessment, and foresight. Although there has been little systematic attention to the conceptual development of the TF field as a whole, the literature profile of the TF field in general shows increasing research activity and interest in TF as the need for TF increases.

There are hundreds of methods being used for TF. Many experts in the field agree that it is advantageous to use several methods simultaneously, as each method can only deal with limited aspects of a forecasting case. The quality of forecasts greatly depends on proper selection and application of appropriate method.

Several studies emphasize that TF in practice, especially in companies, is an unstructured and unsystematic process – which illustrates the opportunity for improvement. Enhanced access to information offers particular promise to improve TF. In an era when tremendous research and development activity worldwide results in explosive growth in the amount of scientific and engineering literature, the MIT/MIST research on developing novel methods for automatically mining science and technology information sources will contribute to this improvement.

3.2 MIT/MIST APPROACH

To facilitate the formation of research strategies with the greatest potential, MIT and MIST started a collaborative research project called "Technological Forecasting using Data Mining and Semantics" (TFDMS). The study focuses on novel methods for automatically mining science and technology information sources with the aim of extracting patterns and trends. The goals include (but are not limited to) generating growth forecasts for technologies of interest, intuitive representations of interrelationships between technology areas, identification of influential researchers or research groups and the discovery of underlying factors, which may affect or stimulate technological growth. The aim is to develop a suite of techniques, which will significantly extend and improve existing methods for performing so-called "tech-mining" [Woon & Madnick 2008], [Woon et al 2009(1)], [Woon et al. 2009(2)].

The work described in this thesis applies bibliometric analysis techniques mentioned in the TF literature above. As a novelty, we automatically generate the terms we analyze by using online information sources without asking experts to come up with them. We then use a hit count trend analysis to create a list of technology areas that are likely to grow exponentially. These lists are presented to experts for further review to complement their analysis.

From the technical perspective, our main focus in this thesis, however, is more on developing a better data access platform to collect the information needed in TF analyses than on advancing an algorithmic aspect of a particular TF technique, which is being done by other researchers in our group. For example, [Woon & Madnick 2008(1)] presents a novel method for automatically constructing taxonomies for specific research domains. The proposed methodology uses term co-occurence frequencies as an indicator of the semantic closeness between terms.

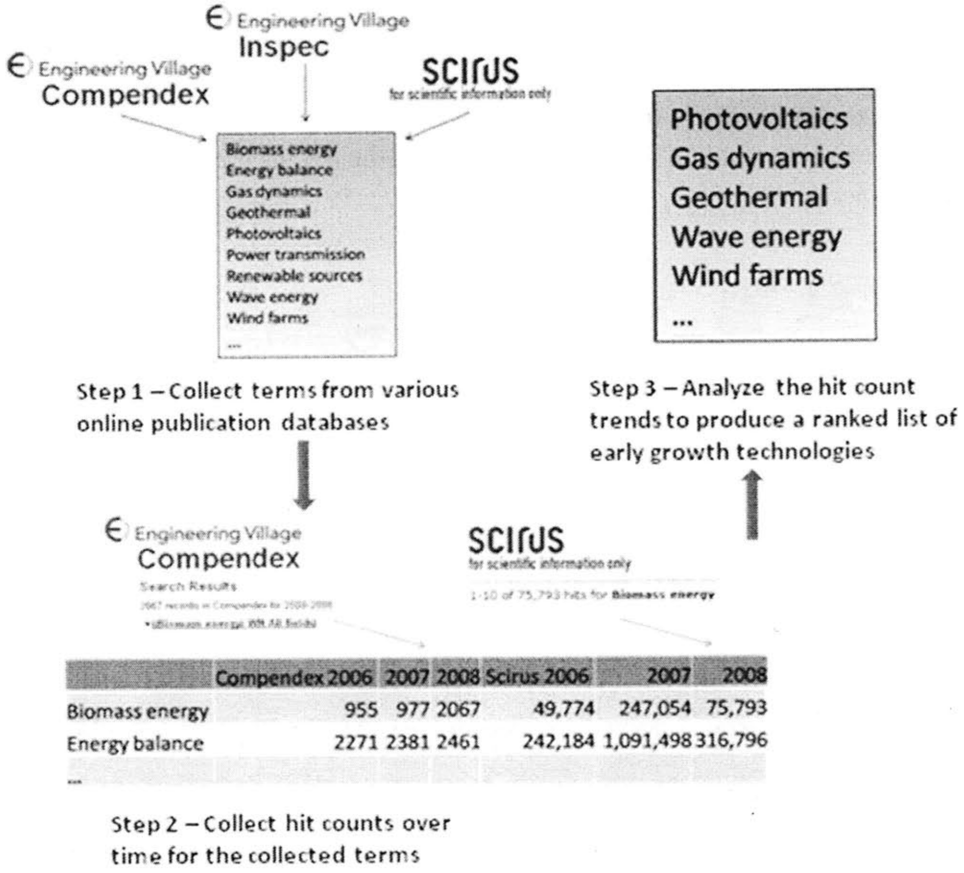
[Ziegler et al. 2008] presents an approach to bibliometric analysis in the context of technology mining using Latent Semantic Analysis (LSA) to reveal the concepts that underlie the terms relevant to a field. [Ziegler 2009] presents methods and software implementation for analyzing a field of research through the use of bibliometrics. [Camiña 2010] investigates the modeling of research landscapes through the automatic generation of hierarchical structures (taxonomies) comprised of terms related to a given research field. Several different taxonomy generation algorithms are discussed and analyzed within this paper, each based on the analysis of a data set of bibliometric information obtained from a credible online publication database.

The set of tools we created for this thesis separate generic and specific aspects of Web data access to make the data collection process as easy and extendible as possible and are discussed in more detail in the following chapters.

CHAPTER 4: EARLY GROWTH TECHNOLOGY ANALYSIS

We mentioned in the introduction that early growth technology analysis consists of three key steps (also illustrated in Figure 4 below):

1. Term collection by extracting related keywords from articles for a given area of interest
2. Determining the occurrence frequencies of these keywords (hit counts)
3. Identifying those exhibiting rapid growth, particularly if starting from a low base


Engineering Village
Inspec
Engineering Village
Compendex
SCIRUS
for scientific information only
Biomass energy
Energy balance
Gas dynamics
Geothermal
Photovoltaics
Power transmission
Renewable sources
Wave energy
Wind farms
...
Photovoltaics
Gas dynamics
Geothermal
Wave energy
Wind farms
...
Step 1 – Collect terms from various online publication databases
Step 3 – Analyze the hit count trends to produce a ranked list of early growth technologies
Engineering Village
Compendex
Search Results
SCIRUS
for scientific information only
1-10 of 75,793 hits for Biomass energy
...
Step 2 – Collect hit counts over time for the collected terms

	Compendex 2006	2007	2008	Scirus 2006	2007	2008
Biomass energy	955	977	2067	49,774	247,054	75,793
Energy balance	2271	2381	2461	242,184	1,091,498	316,796

Figure 4: Three keys steps for Early Growth Technology Analysis

In this chapter, we are going to describe each step in more detail.

4.1 ANALYSIS STEPS

4.1.1 Collection of Relevant Terms

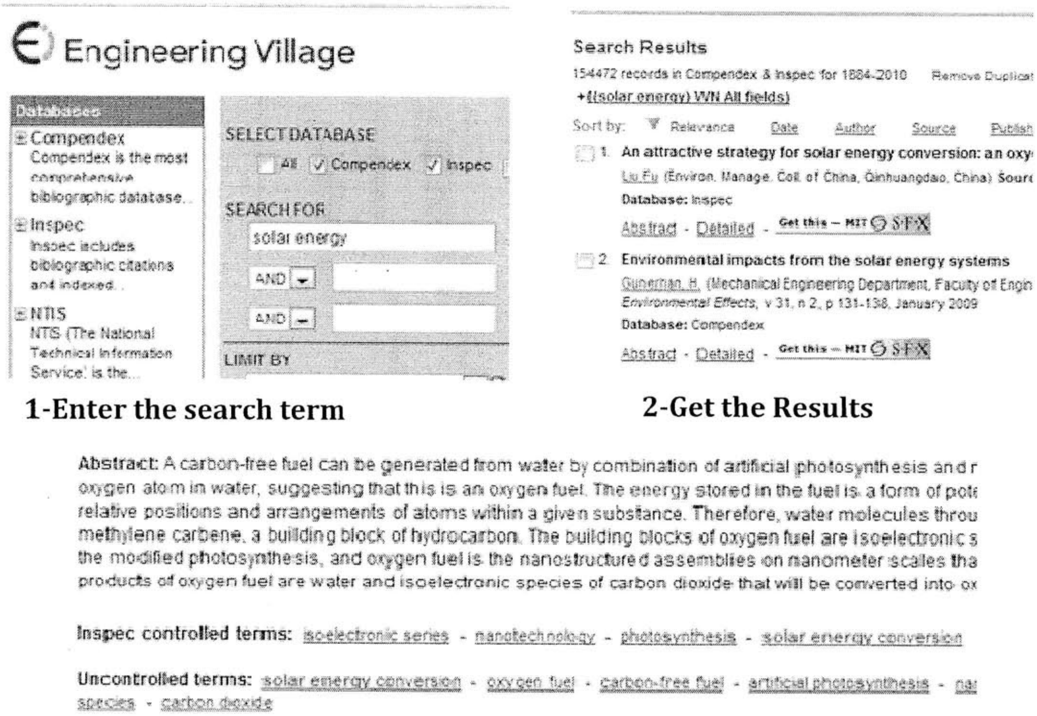
Term collection starts with a seed term such as "renewable energy" that acts as a proxy for the general technology area of interest. We then utilize online publication databases such as Compendex, Inspec, and Scirus to find terms relevant to this seed term. Some of these online databases are freely available (e.g. Scirus), some require subscription (e.g. Compendex and Inspec), and yet others require permission for programmatic access (e.g. Google Scholar).

In order to be useful for term collection, a data source must have a way of returning relevant terms given a seed term. This can be a one step process, when the site simply returns a relevant term list given a query with the seed term, or may involve merging the results of multiple queries that traverse multiple pages. As it will be explained in more detail in the following chapters, we treat these online sources as if they are databases that can respond to queries in SQL (Structured Query Language) via web wrappers. Thus each data source, appropriately wrapped, is accessed uniformly from our software tools. This access uniformity enables us to easily add new data sources into our system.

Below we explain how the Compendex, Inspec and Scirus data sources are used for term collection.

Compendex & Inspec

Compendex and Inspec are data sources provided by Engineering Village, and we can access both sources via the same interface. In Figure 5 below, we illustrate how terms can be collected from these data sources. A search term submitted to the data source returns a number of articles with links to abstracts. When the abstract is accessed, two term lists, controlled and uncontrolled, can be seen. These term lists describe the contents of the article, thus presumably are relevant to our original search term.



Engineering Village

Databases
Compendex
Compendex is the most comprehensive bibliographic database...
Inspec
Inspec includes bibliographic citations and indexed...
NTIS
NTIS (The National Technical Information Service) is the...

SELECT DATABASE

□ All ☑ Compendex ☑ Inspec

SEARCH FOR

solar energy

1-Enter the search term

Search Results

154472 records in Compendex & Inspec for 1884-2010 Remove Duplicat

+((solar energy) WN All fields)

Sort by: Relevance Date Author Source Publish

1. An attractive strategy for solar energy conversion: an oxy
Liu Fu (Environ. Manage. Coll. of China, Qinhuangdao, China) Sourc
Database: Inspec
Abstract - Detailed - Get this – MIT SFX

2. Environmental impacts from the solar energy systems
Gunerhan, H. (Mechanical Engineering Department, Faculty of Engin
Environmental Effects, v 31, n 2, p 131-138, January 2009
Database: Compendex
Abstract - Detailed - Get this – MIT SFX

2-Get the Results

Abstract: A carbon-free fuel can be generated from water by combination of artificial photosynthesis and r
oxygen atom in water, suggesting that this is an oxygen fuel. The energy stored in the fuel is a form of pot
relative positions and arrangements of atoms within a given substance. Therefore, water molecules throu
methylene carbene, a building block of hydrocarbon. The building blocks of oxygen fuel are isoelectronic s
the modified photosynthesis, and oxygen fuel is the nanostructured assemblies on nanometer scales tha
products of oxygen fuel are water and isoelectronic species of carbon dioxide that will be converted into ox

Inspec controlled terms: isoelectronic series - nanotechnology - photosynthesis - solar energy conversion

Uncontrolled terms: solar energy conversion - oxygen fuel - carbon-free fuel - artificial photosynthesis - na
species - carbon dioxide

3-Collect controlled and uncontrolled terms

Figure 5: Term Collection from Compendex & Inspec

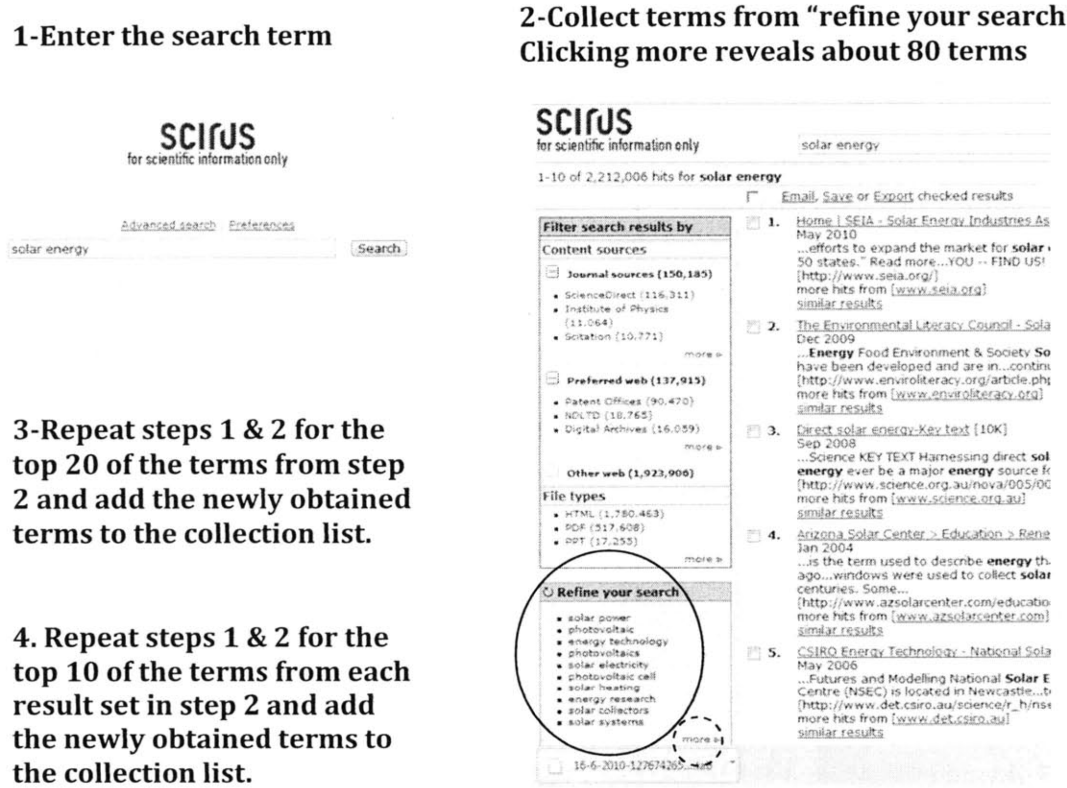
The Engineering Village site makes the distinction between controlled and uncontrolled terms as follows:

"Uncontrolled terms, also known as free language terms, are additional subject terms assigned by indexers. These terms are not selected from the Inspec Thesaurus, but can reflect new expressions and terminology used in a particular discipline. These terms allow for further specificity in indexing that is not available using controlled vocabulary. Uncontrolled terms may subsequently become part of Inspec's controlled vocabulary."

Thus, including the uncontrolled terms to the relevant term list may make the resulting term list more interesting for the analysis by increasing the specificity of the individual terms. Term collection this way can be done until a pre-specified maximum number of abstracts are reached.

Scirus

Collecting terms from the Scirus data source involves a recursive process as illustrated in Figure 6 below. Searching a term in Scirus not only brings links to articles, but also a list of 80 relevant terms that are added to our term list. We then repeat this process for the top 20 terms in our term list, each bringing 80 terms that are added to the term list. At this level we will have about 1680 terms assuming no duplicates. If we want still more terms, we take the top 10 terms of the result-sets returned by the first 20 terms in previous step, and repeat the whole process again. This results in 20*10*80 = 16000 more terms and is often more than enough. The choice of 20 and 10 ensures a good balance of depth vs. breadth in our search process and can be adjusted easily. It is also possible to specify additional levels in the process.



1-Enter the search term

2-Collect terms from "refine your search" Clicking more reveals about 80 terms

3-Repeat steps 1 & 2 for the top 20 of the terms from step 2 and add the newly obtained terms to the collection list.

4. Repeat steps 1 & 2 for the top 10 of the terms from each result set in step 2 and add the newly obtained terms to the collection list.

1-10 of 2,212,006 hits for **solar energy**

1. Home | SEIA - Solar Energy Industries As
May 2010
...efforts to expand the market for **solar**
50 states." Read more...YOU -- FIND US!
[http://www.seia.org/]
more hits from [www.seia.org]
similar results

2. The Environmental Literacy Council - Sola
Dec 2009
...**Energy** Food Environment & Society **So**
have been developed and are in...contin
[http://www.enviroliteracy.org/article.ph
more hits from [www.enviroliteracy.org]
similar results

3. Direct solar energy-Key text [10K]
Sep 2008
...Science KEY TEXT Harnessing direct **sol**
energy ever be a major **energy** source f
[http://www.science.org.au/nova/005/00
more hits from [www.science.org.au]
similar results

4. Arizona Solar Center > Education > Rene
Jan 2004
...is the term used to describe **energy** th
ago...windows were used to collect **solar**
centuries. Some...
[http://www.azsolarcenter.com/educatio
more hits from [www.azsolarcenter.com]
similar results

5. CSIRO Energy Technology - National Sola
May 2006
...Futures and Modelling National **Solar E**
Centre (NSEC) is located in Newcastle...t
[http://www.det.csiro.au/science/r_h/ns
more hits from [www.det.csiro.au]
similar results

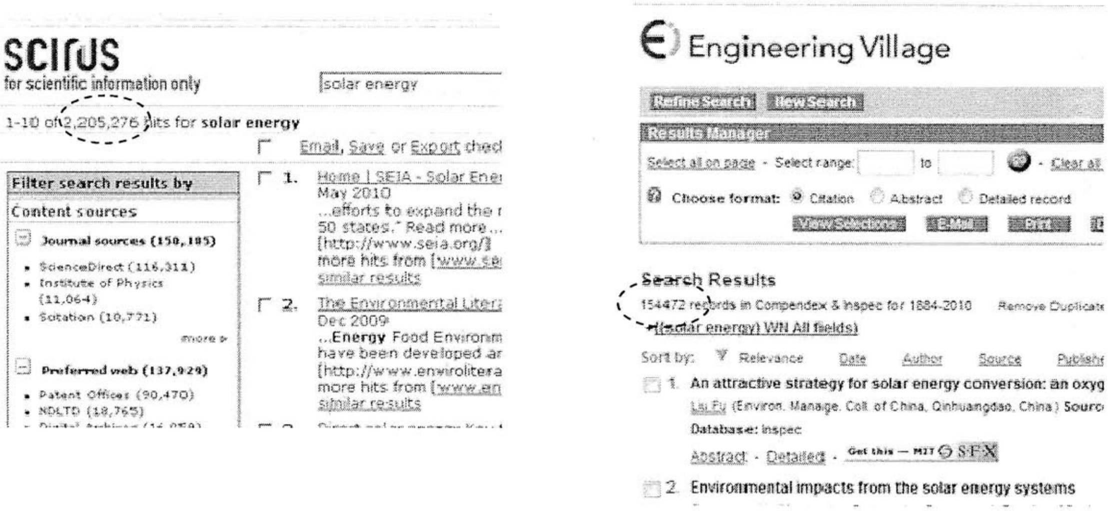
Figure 6: Term Collection from Scirus

During term collection, we eliminate duplicates and provide the option to combine results from different databases together. Our observation is that there is not much overlap between these three databases, and merging results may offer a richer term list.

In some cases, term collection can be polluted by irrelevant terms. In those situations, we offer a refinement strategy that allows terms to be added to the list only if a term's direct relevancy term list contains our original seed term. For example, for the seed term solar power, "photovoltaics" is returned as a relevant term in the first 80 results. To ascertain it is indeed a relevant term, we collect a new term list for the seed term photovoltaics and check that solar power exists in the first 80 terms of its relevancy term list. While this feature is quite helpful in cutting the number of irrelevant terms, it also limits the diversity of our list. Another option to reduce irrelevant terms is to add the seed term to the recursive searches. For example, instead of searching for the term list of "photovoltaics", we can search the term list of "solar power photovoltaics". The trade-off between relevancy and diversity also exists in this technique.

4.1.2 Collection of Hitcounts

After collecting the terms from a set of sources, our next step is to find out hit counts of each term over a number of years. For this task, we use the hit counts returned by a set of data sources, which in our case are the same sources we use for collecting terms. Note that these two tasks are separate and the data sources need not be same. In Figure 7 below, we show where we get our hit counts in Scirus, Compendex and Inspec data sources. Note that Engineering Village has a uniform interface for accessing hitcounts from Compendex and Inspec data sources. It is possible to get hitcounts separately from each source, or in combined form. These data sources provide hitcounts for a time interval, and by narrowing these intervals to single years, we can extract hitcounts for a particular year.



solar energy

1-10 of 2,205,276 hits for **solar energy**

□ Email, Save or Export chec

Filter search results by

Content sources

Journal sources (150,185)
- ScienceDirect (116,311)
- Institute of Physics (11,064)
- Scitation (10,771)

more »

Preferred web (137,929)
- Patent Offices (90,470)
- NDLTD (18,765)

□ 1. Home | SEIA - Solar Ene
May 2010
...efforts to expand the r
50 states." Read more...
[http://www.seia.org/]
more hits from [www.se
similar results

□ 2. The Environmental Liter
Dec 2009
...**Energy** Food Environm
have been developed ar
[http://www.envirolitera
more hits from [www.en
similar results

Refine Search · New Search

Results Manager

Select all on page - Select range: to - Clear all

Choose format: ● Citation ○ Abstract ○ Detailed record

View Selections · E-Mail · Print

Search Results

154472 records in Compendex & Inspec for 1884-2010 Remove Duplicate

((solar energy) WN All fields)

Sort by: Relevance Date Author Source Publishe

□ 1. **An attractive strategy for solar energy conversion: an oxyg**
Lai Fu (Environ. Manage. Coll. of China, Qinhuangdao, China) Sourc
Database: Inspec
Abstract - Detailed - Get this — MIT SFX

□ 2. **Environmental impacts from the solar energy systems**

Figure 7: Hitcount Collection from Scirus, Compendex & Inspec

Hit count collection is a slow process, because it involves fetching pages over the network. For a term list size of 5000, and a time span of three years, we would need to send 15000 queries to the databases. Although we use parallel execution, the sites accept limited number of concurrent queries at a time, so we have to do much of the task sequentially. It takes about 2 hours to complete the 15000 queries mentioned above. In section 4.4 we explain the details of extracting data from websites.

4.1.3 Ranking Terms According To Their Hitcounts

Our goal in this step is to narrow down our original list to a subset of terms that seem to fit the early growth description. Although there are many different ways to identify an early growth technology in terms of its hit count trend, we currently rank each term according to the percentage increase of its hitcounts from initial to final year, or by looking at the log ratio of final year hit counts to initial year hitcounts. More specifically both formulas are defined as follows:

Formula 1(Early Growth) = Log(hits in End Year) / Log(hits in Beginning Year)
Formula 2(Total growth) = (hits in End Year – hits in Beginning Year)/hits in Beginning Year

By ranking terms according to the log ratio of final hit count to initial hit count, we favor terms with low initial hit counts. For example, a term whose hit counts increased from 1 to 3 will be ranked higher than a term whose hit counts increased from 1000 to 2000 over the same time period. This is in line with our interest in identifying technologies that are not yet well-known, but are on the steepest part of their growth curve.

Both of these formulas need an adjustment when the initial year hitcounts is zero to avoid division by zero. In that case we use a small number such as 0.0001 to avoid this error, and also assign significance to this rare event. The choice of this number is unfortunately somewhat arbitrary, but the smaller it is defined to be, the higher the ranking zero-hit count terms will receive.

As this research progresses, other metrics besides those described above will be added. Our goal at this point is to create an automated and extendible approach to technology forecasting. When new formulas are desired, the system need to be easy to extend to accommodate the new approaches.

4.2 IMPLEMENTATION APPROACH

As explained in the steps above, the majority of our work is about extraction of data from web sites of interest. In our research group, researchers have used three different approaches to accomplish this task. These approaches can be characterized as follows:

(1) BLACK-BOX: Developing a program from scratch.

(2) MODULAR: Developing a parameterizable program with generic and web-site specific modules.

(3) DECLARATIVE: Developing a customable package that hides the generic program logic, and only exposes the declaratively specified web-site specific logic.

We chose to follow approach three, and believe that it is superior to other two approaches. With the hope of aiding future researchers, who may undertake the web data extraction task in this or other domains, we present a comparison of these three methods along the following dimensions:

1) Authoring: The process of creating a wrapper for a web source.

2) Maintainability: The process of updating an existing wrapper.

3) Teachability: The process of teaching wrapper development to a new comer.

4) Capability: The power and flexibility of the wrappers.

Although we try to conduct the comparison in an abstract way, from time to time we will refer to actual programs that represent these approaches especially in the capability comparison part. First, we provide some background on the three approaches.

4.2.1 Background

Black-Box Approach

This approach was used by John Baker [Baker 2008] with the objective of leveraging pre-existing wrapper software to meet our project needs without too much effort. Specifically, John used an open source Perl tool called WebSearch.pl which leverages the back-end WWW::Search Perl libraries. With these tools he was able to submit queries against the following search engines on the right. Unfortunately, none of these files correspond to the publication databases online.

These ".pm" files already existed on the web, and each corresponds to a large program dedicated to a single web site. In order to wrap other web sources of interest, similar programs needed to be devised from scratch. If the user wanted a specific output format, the .pm file must be modified. An example output of the AltaVista wrapper, only outputting 5 URLs, and in verbose mode gives a title and description as shown below:

- AltaVista.pm
- CraigsList.pm
- Crawler.pm
- Excite/News.pm
- ExciteForWebServers.pm
- Fireball.pm
- FolioViews.pm
- Gopher.pm
- HotFiles.pm
- Livelink.pm
- MetaCrawler.pm
- Metapedia.pm
- MSIndexServer.pm
- NetFind.pm
- Newturfers.pm
- PLweb.pm
- Profusion.pm
- Search97.pm
- SFgate.pm
- Timezone.pm
- Verity.pm

<u>Request</u>:

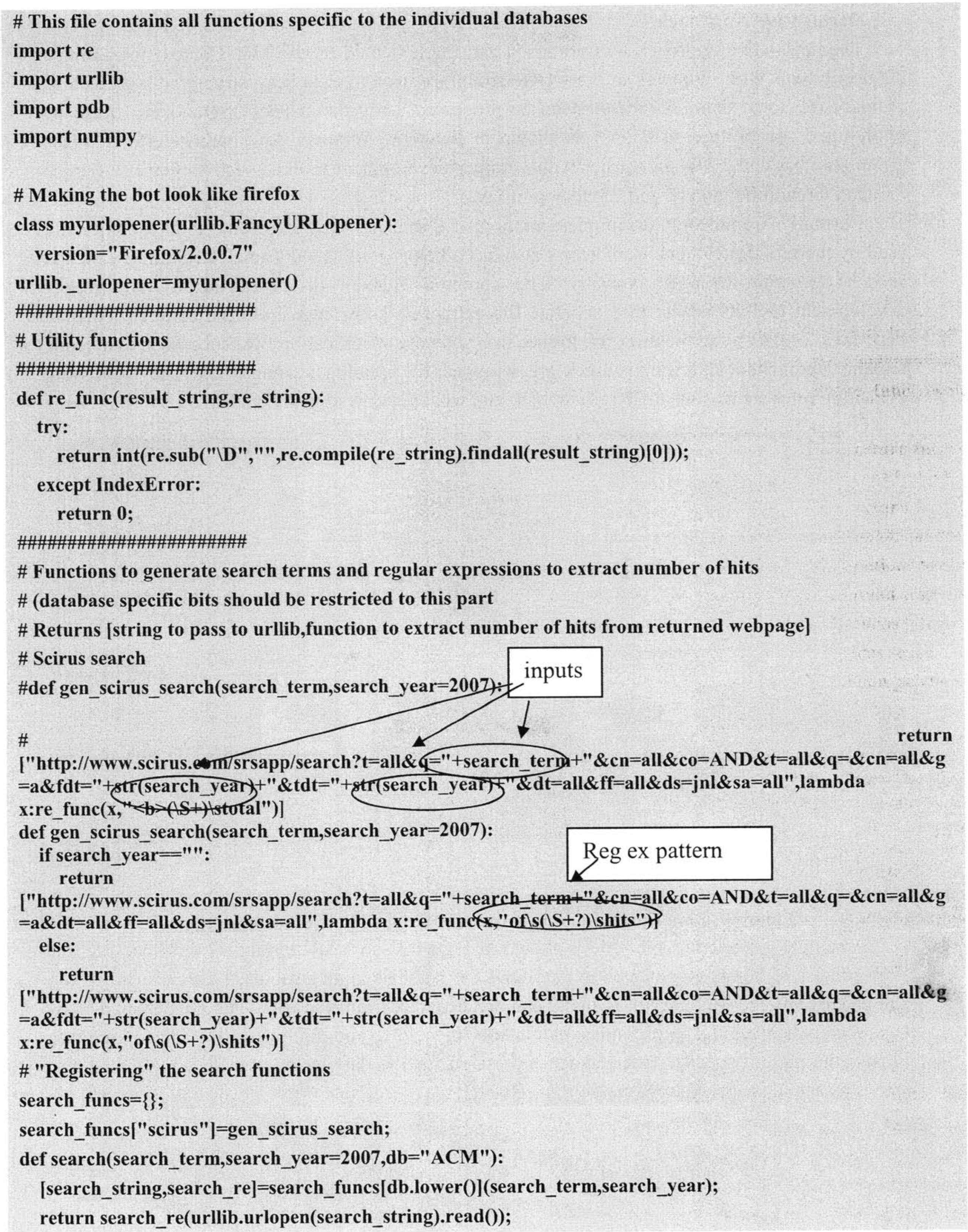
```
  jab@jabTab][17:41 GMT]=> WebSearch -e AltaVista -m 5 "alternate energy" -verbose
```

<u>Output</u>:

```
  2. (title: Alternate Energy Resource Network,
      description: We provide the latest information about alternative energy,
  solar energy and fuel cells with daily updated industry news, articles and
  renewable energy resources)
          http://www.alternate-energy.net/
  3. (title: Alternate Energy Solutions,
      description: Alternate Energy Solutions Inc. Energy Solutions, Which Work ...
  photovoltaic modules and alternate energy solutions powering industrial and ...)
```

Modular Approach

Prof. Woon [Woon & Madnick 2008(2)] took this approach and developed wrappers for Google Scholar, IngentaConnect, Scirus, ACM Guide, SpringerLink and IEEExplore using Python. Unlike the black-box approach, generic functions of data extraction are separated from the web-site specific logic in this approach. When one needs to wrap new sites, these generic utility functions are copied and only the web-site specific parts of the code are modified. Each wrapper is implemented as a program dedicated to a single web site. An example wrapper in Python is shown in Figure 8 for Scirus.



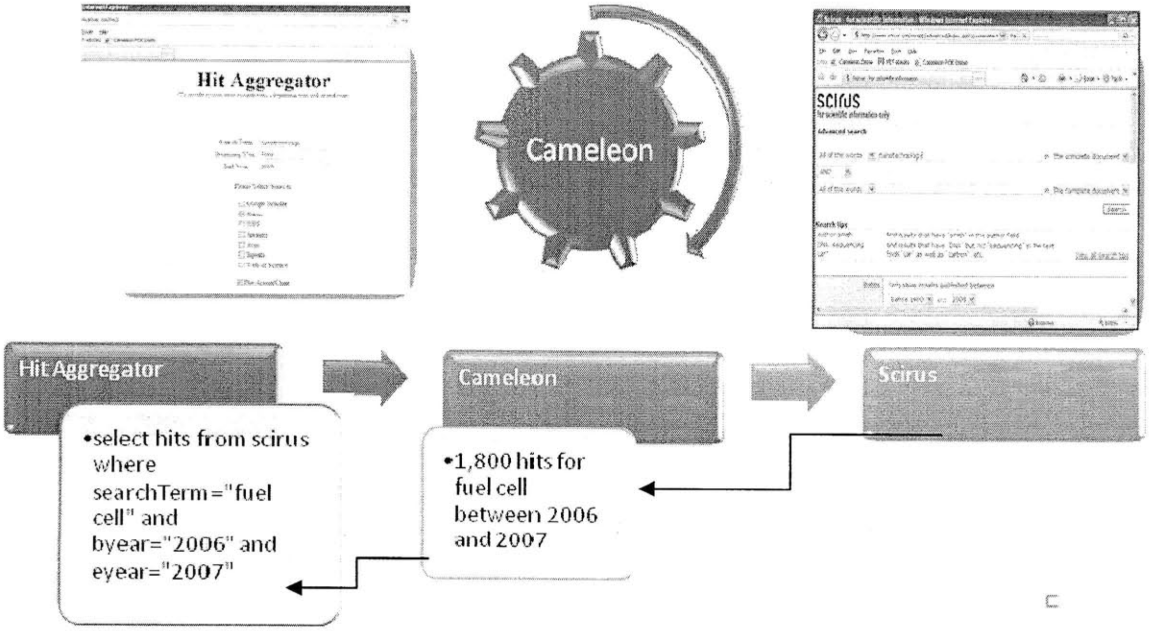
```
# This file contains all functions specific to the individual databases
import re
import urllib
import pdb
import numpy

# Making the bot look like firefox
class myurlopener(urllib.FancyURLopener):
  version="Firefox/2.0.0.7"
urllib._urlopener=myurlopener()
#########################
# Utility functions
#########################
def re_func(result_string,re_string):
  try:
    return int(re.sub("\D","",re.compile(re_string).findall(result_string)[0]));
  except IndexError:
    return 0;
########################
# Functions to generate search terms and regular expressions to extract number of hits
# (database specific bits should be restricted to this part
# Returns [string to pass to urllib,function to extract number of hits from returned webpage]
# Scirus search
#def gen_scirus_search(search_term,search_year=2007):

#                                                                                    return
["http://www.scirus.com/srsapp/search?t=all&q="+search_term+"&cn=all&co=AND&t=all&q=&cn=all&g
=a&fdt="+str(search_year)+"&tdt="+str(search_year)+"&dt=all&ff=all&ds=jnl&sa=all",lambda
x:re_func(x,"<b>(\S+)\stotal")]
def gen_scirus_search(search_term,search_year=2007):
  if search_year=="":
    return
["http://www.scirus.com/srsapp/search?t=all&q="+search_term+"&cn=all&co=AND&t=all&q=&cn=all&g
=a&dt=all&ff=all&ds=jnl&sa=all",lambda x:re_func(x,"of\s(\S+?)\shits")]
  else:
    return
["http://www.scirus.com/srsapp/search?t=all&q="+search_term+"&cn=all&co=AND&t=all&q=&cn=all&g
=a&fdt="+str(search_year)+"&tdt="+str(search_year)+"&dt=all&ff=all&ds=jnl&sa=all",lambda
x:re_func(x,"of\s(\S+?)\shits")]
# "Registering" the search functions
search_funcs={};
search_funcs["scirus"]=gen_scirus_search;
def search(search_term,search_year=2007,db="ACM"):
  [search_string,search_re]=search_funcs[db.lower()](search_term,search_year);
  return search_re(urllib.urlopen(search_string).read());
```

Figure 8: An example wrapper in Python is shown for the Scirus online database.

Declarative Approach

The declarative approach was proposed and implemented by the MIT Context Interchange (COIN) group. Web wrappers in the COIN group are used to treat semi-structured web data as ordinary relational data sources that can be processed using the standard SQL query language (with some capability restrictions) as shown in Figure 9. Wrapper development efforts in the group date back to 1995 and earlier with wrapper development toolkits such as Generic Screen Scraper, Grenouille in Perl, and Cameleon in Java.

Currently, Cameleon#, a reimplementation of Cameleon in .NET, is the toolkit commonly used by the group members. Cameleon# also has a helper tool called Cameleon Studio, which is used to generate wrappers visually. The common element of all of the COIN wrapper development toolkits is that they separate the extraction knowledge from the code (whether in Perl, Java, or C#) by expressing the former in a separate specification file (spec file). The code remains untouched and web sources are wrapped by creating a simple text file expressing extraction rules such as what URL(s) to visit, and what patterns to apply, and so on.

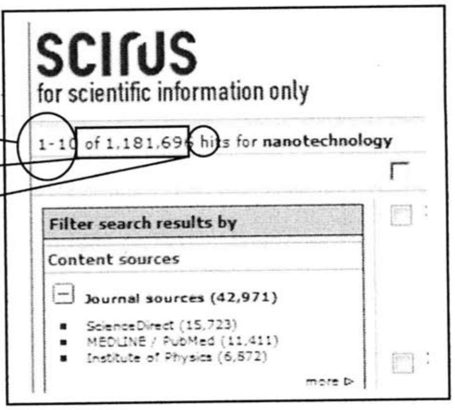


•select hits from scirus where searchTerm="fuel cell" and byear="2006" and eyear="2007"

Figure 9: Simple SQL query against the wrapped Scirus Search Engine.

A sample Cameleon# spec file is shown in Figure 10 in XML format. In this spec file, the Web address of Scirus is indicated in the SOURCE tag. The input attributes (searchTerm, bYear, and eYear) are enclosed within # signs, and are expected from the user. The BEGIN and END tags specify (in regular expressions) landmarks preceding and following the data of interest. Finally the pattern specifies the regular expression for the data to be extracted. Figure 10 also shows an actual snapshot from the Scirus web site.

```
<?xml version="1.0" encoding="UTF-8"?>
<RELATION name="scirus">
 <SOURCE
URI="http://www.scirus.com/srsapp/search?sort=0&amp;t=all&amp;q=#searchTerm#&amp;
cn=all&amp;co=AND&amp;t=all&amp;q=&amp;cn=all&amp;g=a&amp;fdt=#byear#&amp;
tdt=#eyear#&amp;dt=all&amp;ff=all&amp;ds=jnl&amp;ds=nom&amp;ds=web&amp;sa=all"
>
   <ATTRIBUTE name="hits" type="string">
    <BEGIN><![CDATA[1-10]]></BEGIN>
    <PATTERN><![CDATA[of\s(.+?)\s]]></PATT
    <END><![CDATA[hits]]></END>
   </ATTRIBUTE>
  </SOURCE>
 </RELATION>
```



scirus

for scientific information only

1-10 of 1,181,696 hits for **nanotechnology**

Filter search results by

Content sources

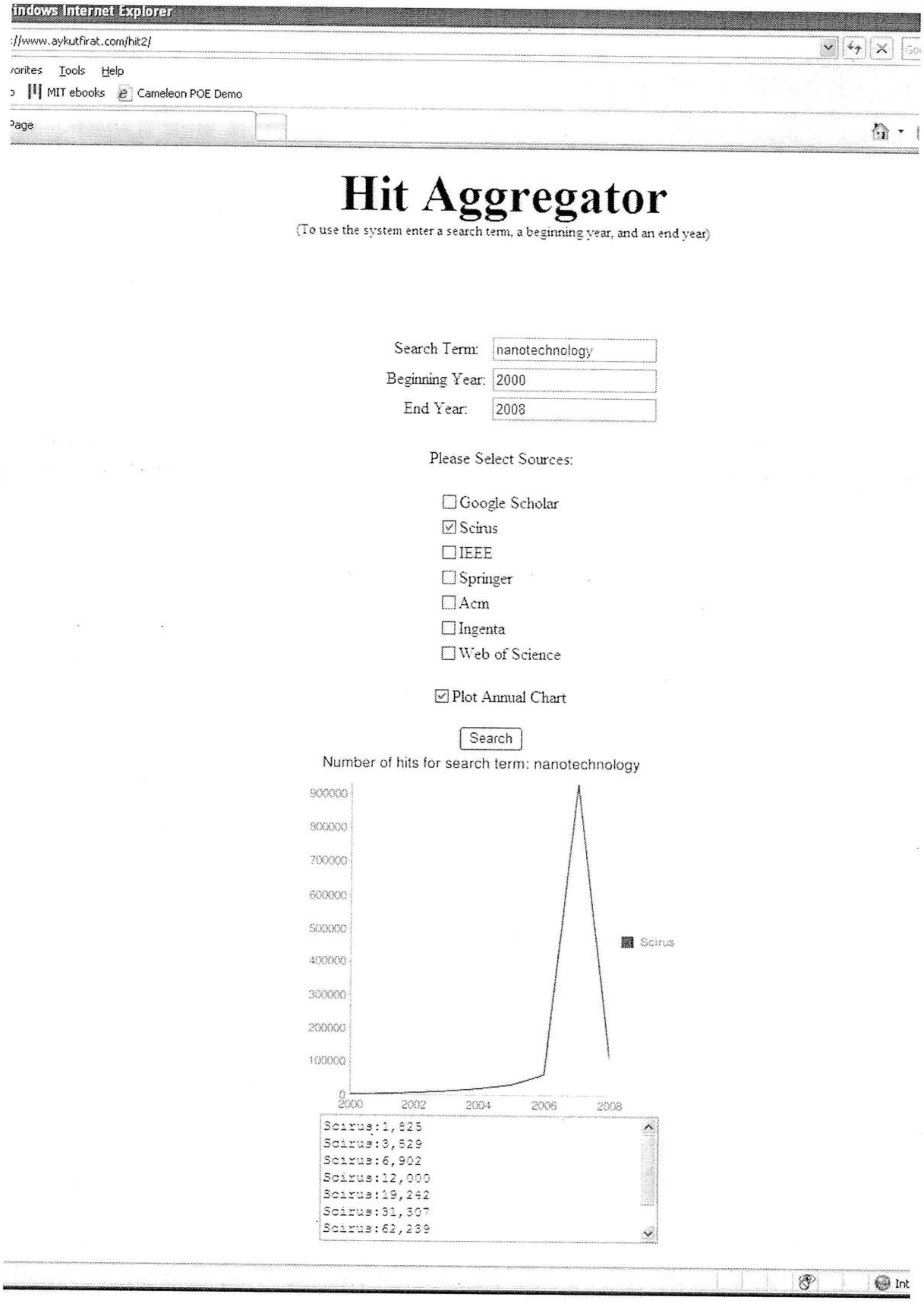
⊟ **Journal sources (42,971)**

- ScienceDirect (15,723)
- MEDLINE / PubMed (11,411)
- Institute of Physics (6,572)

more ▷

Figure 10: Cameleon# Spec File for Scirus Database

The Cameleon data, then can be used in application programs such as the Hit Aggregator we developed for our technology forecasting project as shown in Figure 11.



Hit Aggregator

(To use the system enter a search term, a beginning year, and an end year)

Search Term: nanotechnology

Beginning Year: 2000

End Year: 2008

Please Select Sources:

□ Google Scholar

☑ Scirus

□ IEEE

□ Springer

□ Acm

□ Ingenta

□ Web of Science

☑ Plot Annual Chart

Search

Number of hits for search term: nanotechnology

```
Scirus:1,825
Scirus:3,529
Scirus:6,902
Scirus:12,000
Scirus:19,242
Scirus:31,307
Scirus:62,239
```

Figure 11: Hit Aggregator we developed for our technology forecasting project.

4.2.2 Comparison

A. Authoring

Authoring wrappers for "difficult pages"

Developing a wrapper from scratch or using a modular approach requires basic knowledge of programming languages, regular expressions, and intimate knowledge of its web related libraries. Wrapping is relatively straight forward, if the data resides in a single page that can be accessed with a static and standard URL. When the developer needs to deal with "a difficult page" involving cookie handling, redirects, form submission, SSL, Javascript interpretation, and passing data from one web page to another, even more code, libraries and external programs are needed. Although these libraries and external programs are available, the final Python/Perl program to wrap a "difficult page" will be complex. Under the declarative approach, the wrapper developer need not know any programming language. With the help of visual tools such as Cameleon Studio, the wrapper developer only needs to learn regular expressions, and the structure of a specification file.

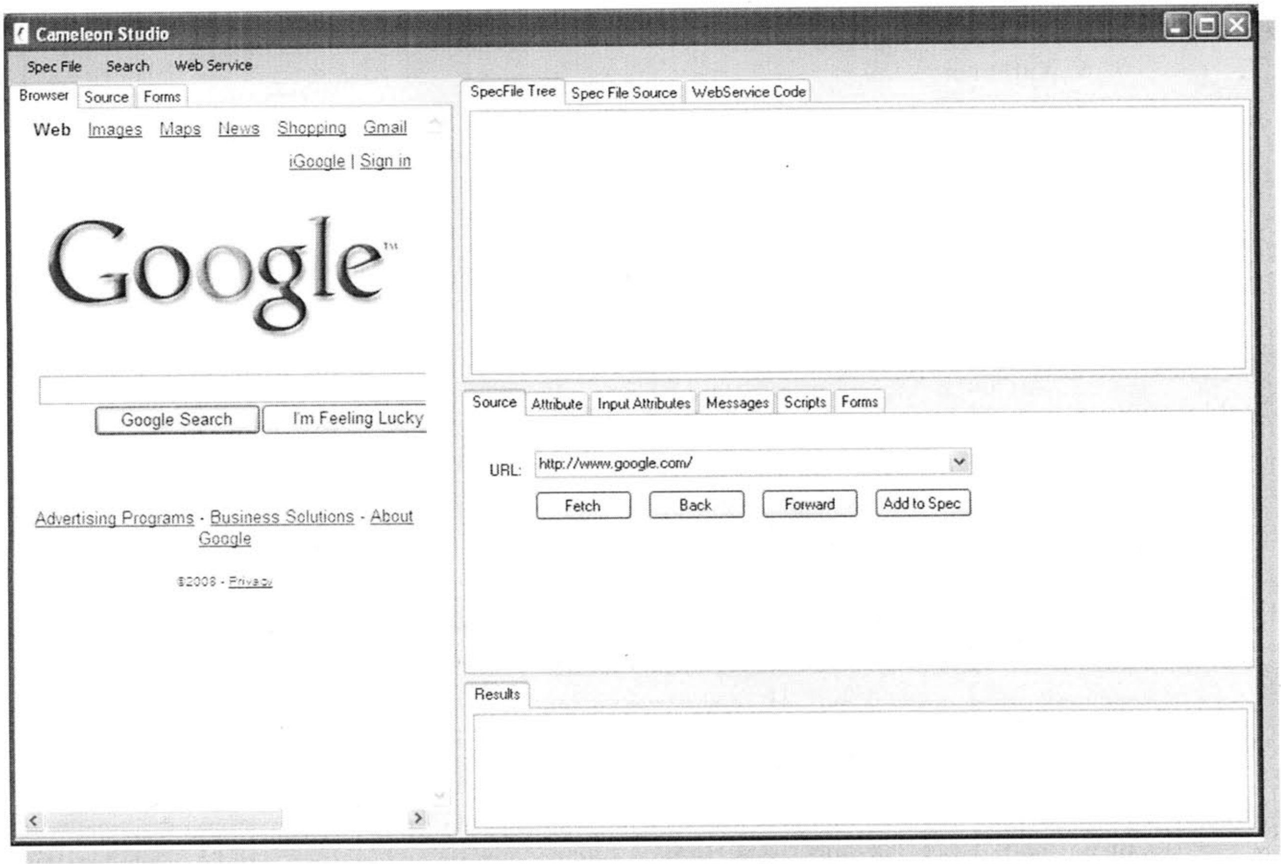
Visual support

One of the most time-consuming aspects of web wrapping is the identification of form elements manually when the target page requires form submission. Without visual tools, the developer using a black-box or modular approach needs to use a text editor to identify form elements manually, and potentially introduce errors into the wrapper code. Using Cameleon Studio, a visual application that aids the development of Cameleon spec files, and converting them into web services, this process is more automatic and forms can be added to the spec file with a single click. Furthermore, with a built-in browser Cameleon Studio provides visual support for identifying landmark text and patterns easily.

As shown in Figure 12, Cameleon Studio has a built in browser on its left that also shows the source of a web page, and the forms that are in that web page. On its upper right it shows the spec file in tree form, and original form, and the auto produced web service code. On the lower right, it has several tabs for surfing web sites (Sources), defining attributes (Attributes), providing values for input attributes (Input Attributes), displaying messages from the program (Messages), displaying scripts from the web sites, and authoring custom forms (forms). Test results can also be viewed via the Results tab.

Special-purpose debugging

The wrapper developer may sometimes encounter errors when trying to wrap web pages. When the wrapper is developed using Python/Perl, the user is limited to the debugging support of the programming environment. Users of Cameleon Studio, however, are provided with special purpose debugging support. They can, for instance, visually find out what text the regular expression patterns match with a single click, and observe a simulated run of the specification file. This special-purpose debugging support is an important element in speeding up the wrapper development.



URL: http://www.google.com/

Figure 12: Cameleon Studio Interface.

B. Maintenance

One of the fragile aspects of wrapper development is the autonomy of web pages, and their tendency to change their page structure frequently. Wrappers, therefore, need to be updated when the patterns no longer match the desired information in the page. Maintainability of wrappers thus needs to be considered in comparing wrapper development approaches.

Object-Oriented Design Principle: Encapsulate what varies

One of the well-known principles of object oriented software development is separating parts of the code that varies from the parts that stay the same. The primary consequence of applying this principle is better maintainability. In wrapper development, parts that vary from one wrapper to another are extraction knowledge, such as URLs, patterns, form elements, and so on. In Cameleon#, the declarative approach, these varying parts have been separated from the core code that stays the same for all wrappers. This design enables superior maintenance, because updates are limited to the specification files. Developers cannot introduce any unintended errors to the core hidden code.

This is not the case when both the code and extraction knowledge are lumped together in adopting a black-box approach. Modular approach is clearly superior to black-box approach, because changes will be limited to the web-specific parts of the code. In either case, however, there is the danger of wrongly modifying part of the code and making the whole process more time-consuming and error-prone. The convolution of code and knowledge hinders automatic maintenance approaches as well. It is hard to automate maintenance, when there is no fixed structure in a wrapper file. When the extraction knowledge is separated from the code, however, automatic maintenance approaches can be devised utilizing the well-defined structure of the specification file.

C. Teachability

A wrapper development approach will not gain acceptance if it is not easy to teach and learn. One important criticism against popular programming languages used in wrapper development like Perl is that it allows its developers to author obfuscated code more easily than other languages. The following code, which uses a rendition of rotated 90 degree Mayan numerals to extract the required text, from the 2000 Obfuscated Perl Contest is a testimony to the obscurity potential of this language.

```
#:: ::-| ::-| .-. :||-:: 0-| .-| ::||-| .:|-. :||
open(Q,$0);while(<Q>){if(/^#(.*)$/){for(split('',$1)){$q=0;for(split){s/\|
/:.:/xg;s/:/../g;$Q=$_?length:$_;$q+=$q?$Q:$Q*20;}print chr($q);}}}print"\n";
#.: ::||-| .||-| :|||-| ::||-| ||-:: :|||-| .:|
```

While Python is a more readable language than Perl, it does not fare well in terms of readability when compared to the specification files used in Cameleon#. What is more, specification files are displayed in a tree like structure in Cameleon Studio making it even easier to understand. For novice users, learning Cameleon would be much faster than learning a full programming language.

D. Capability

Flexibility

Capabilities of a declarative wrapper engine like Cameleon# are pre-defined and its code is closed to modification as far as the wrapper developer is concerned. New versions of the code can be released by Cameleon# developers, and its capabilities can be expanded, but the wrapper developer is not expected to undertake this task. This is not the case when developing a wrapper using a programming language such as Python/Perl under the black-box or modular approaches. Code is open to modification all the time; therefore the developer has the full flexibility of a programming language. If a page with unforeseen intricacies is encountered, the Python/Perl developer can find a way to overcome the problem. For example, certain types of Google search results are divided in multiple pages, and Cameleon# currently does not offer an easy way to wrap results dispersed over an unknown number of pages. This is, however, quite easy to do in Python/Perl by using a loop. Another example can be given concerning the use of session IDs. While session IDs are extracted from the web page itself when wrapping with Cameleon#, there are no mechanisms to auto generate these Session IDs. A python based wrapper, on the other hand, can easily embed a function to generate legitimate session IDs.

SQL Interface

Cameleon# primary interface accepts simple SQL queries and returns results in table or XML formats. This has several important advantages: (1) Many programmers are familiar with SQL, so writing requests to extract data from web sites is easy to do (assuming that the spec file has already been created for that web site and (2) there are many software systems and tools that have been developed that use the SQL interface (The Excel spreadsheet software is such an example and explained in the next section.) Wrappers developed in Python/Perl usually have custom-designed interfaces (usually simpler and more limited than SQL) or would need to implement similar SQL communication patterns, which would add considerable complexity, in order to be compatible with existing systems.

Excel Integration

Microsoft Excel 2007 allows users to retrieve data from the Web and databases in several ways. By using a web query file, we can include SQL queries directed to web sites (which have Cameleon# spec files) and import the Cameleon# results into Excel. This enables Cameleon# users to easily create elaborate Excel based applications similar to the Hit Aggregator (see Appendix 2). As Excel is frequently used in business settings and is very familiar to many people who do not view themselves as "programmers," being able to use simple SQL queries to selectively import web data into Excel is an important advantage. Wrappers developed in Python/Perl need to function as a server, and be able to return results in HTML/XML format in order to replicate this capability.

Error Handling

When a page changes, or something goes wrong during the wrapping process, elegant error handling becomes important. Cameleon#, unfortunately, does not have such good error handling routines. Errors like "Specified cast is invalid" does not tell much about what went wrong. The user needs to go through the debugging process to get an idea about the error. Perl and Python based approaches can have custom error handling as each wrapper is a program by itself, but this, of course, needs to be programmed. A typical hastily written wrapper program will not have much error handling either.

Java Script Interpretation

Some web pages are based on Javascript, and sometimes the data of interest may be generated during run time via the execution of some Javascript on the web page. In those cases the wrapper needs to be capable of interpreting JavaScript code and utilizing its output. Cameleon# has already some built in JavaScript interpretation support, but it has not been used in many cases. The code takes advantage of .NET framework's ability to mix languages, and can be extended easily. The Perl and Python based wrapper approaches would need to utilize JavaScript libraries that are being made available by the larger Perl and Python community to accomplish the same task.

Summary

For our purposes, wrapper development using Cameleon# is a better approach than either the black-box or modular approaches when we consider the ease of authoring, maintainability, and Teachability, even though both Python and Perl based modular programming approaches are more flexible especially in dealing with pages that are idiosyncratic and cannot be currently handled by Cameleon#. (A summary of the comparison of between Cameleon# and Python/Perl programming approach can be found in Table 6.)

Because we would like to develop a user-friendly tool which can be used by executives at IBM/Masdar, we need to use an approach that is easily maintainable and extensible. For that reason, we believe that the declarative approach to data collection using Cameleon# would best fit this set of requirements. In the next chapter we describe the implementation of a user friendly EGTA tool that builds upon this declarative data collection approach.

	Black-Box	Modular	Declarative
Authoring	Code and knowledge is lumped together and exposed.	Code and knowledge is separated, but code is not hidden from the user.	Code and knowledge are separated, code is hidden, and knowledge is declaratively specified. The wrapper developer need not know any programming language. With the help of Cameleon Studio, the wrapper developer only needs to learn regular expressions, and the grammar of a specification file.
	Developing a wrapper in Python or Perl requires basic knowledge of these programming languages, regular expressions, and knowledge of its web related libraries. To deal with "a difficult page" involving cookie handling, redirects, form submission, SSL, Javascript interpretation, and passing data from one web page to another, even more code, libraries and external programs are needed.		
Maintanence	Updates to code and extraction knowledge are needed.	Updates are limited to the web-specific parts of the code.	Updates are limited to the declarative specification files.
	Automatic maintenance approaches would be difficult, if not impossible, to implement.		Automatic maintenance approaches can be devised utilizing the well-defined structure of the specification file.
Teachability	Popular web wrapping languages such as Perl/Python allows its developers to author obfuscated code more than other languages. They are hard to understand and teach		Specification files are displayed in a tree like structure in Cameleon Studio making it easier to understand. For novice users, it is clear that teaching Cameleon would be much faster than teaching fledged real programming language.
Capability	Code is open to modification, therefore the developer has the full flexibility of a programming language.		Capabilities are pre-defined and closed to modification until a new version is released.

Table 6: A summary of the comparison of between black-box, modular, and declarative approaches

CHAPTER 5: EARLY GROWTH ANALYSIS SOFTWARE TOOLS

In this section, we describe the software tools we developed to support our approach to technology forecasting described in this Thesis. All of our tools use Cameleon Web Wrapper engine [Firat et al. 2005] to query web sources as if they are structured databases using SQL. We started with a simple tool that showed the trend of a technology term over the years from selected scientific databases, before finally developing a much more sophisticated tool that identifies technologies with high growth potential from a seed term.

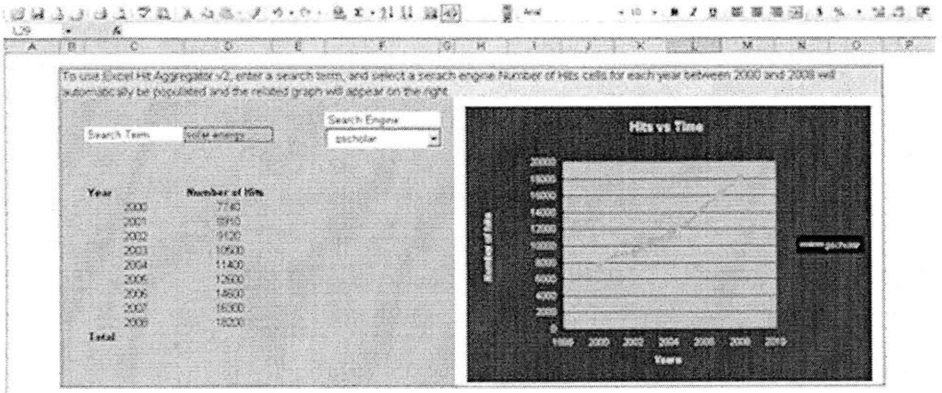
The first tool we created, called hit aggregator, collects and graphs the "hit counts" of terms obtained from different sources over a specified time interval. Hit counts refer to the number of results relevant for a term in a scientific database such as Google Scholar. With this tool we are able to visualize the popularity of technological terms in scientific databases. This tool has Web and Excel versions (see Figures 13 and 14).

The second tool we developed, called Cameleon Scheduler, is used to schedule Cameleon queries to run automatically at periodic intervals between specified dates. The tool is generic and can be used to run any Cameleon query. We use it to collect hit counts of terms provided as a list. The tool also has the capability to automatically build Cartesian product of terms using columns in a given file.

Finally, we created the Early Growth Technology Analysis (EGTA) tool that starts with a seed term to create a term list, collect the hit count of each term over time, and apply a formula to pinpoint the most promising terms. This tool has both desktop and Web versions. Below, we provide an overview of these tools leaving most of the technical details in the Appendices.

5.1 HIT AGGREGATOR

The Hit Aggregator allows users to see how the hit counts trend over the years in different scientific databases. As shown in Figure 14, we use Google Scholar, Scirus, IEEE, Springer, ACM, Ingenta, and Web of Science as our primary scientific databases. The user selects one or many of these databases, enters a term to act as a proxy for a technology, and specify the time interval before the graph of the hit counts are shown. We also created an Excel version of this tool by taking advantage of the Excel integration of the underlying web wrapper engine Cameleon. The details of the Excel version and how it can be set up can be found in Appendix 2.



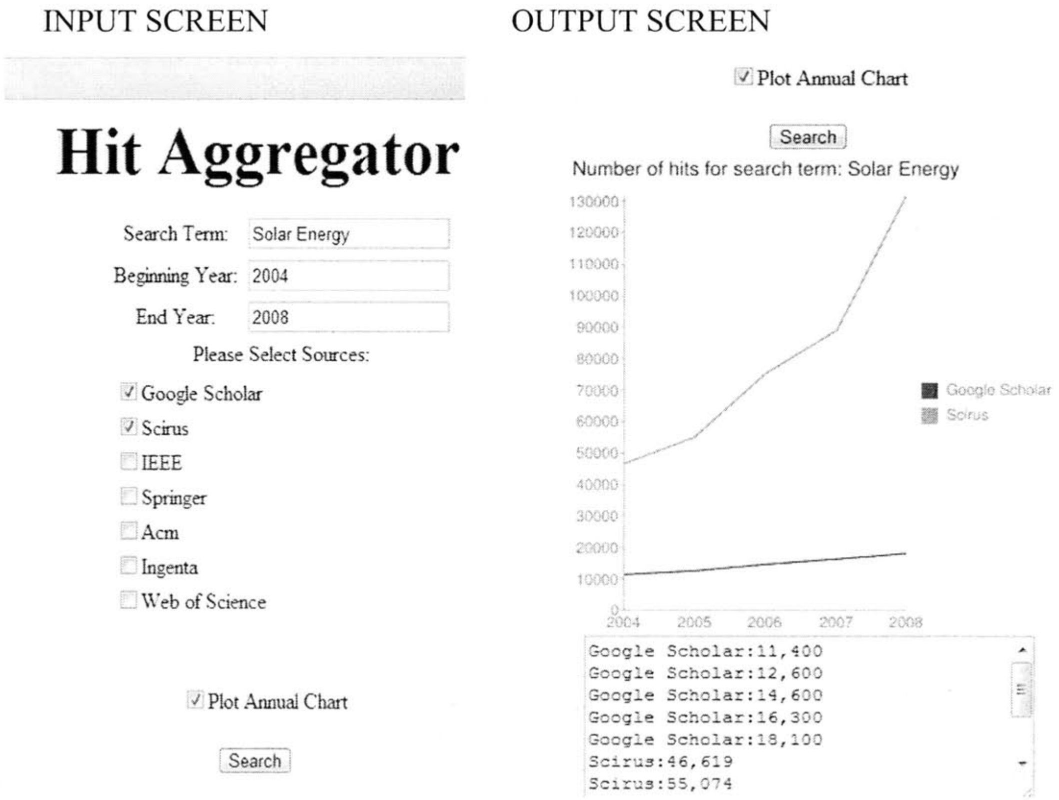
To use Excel Hit Aggregator v2, enter a search term, and select a serach engine Number of Hits cells for each year between 2000 and 2008 will automatically be populated and the related graph will appear on the right

Search Term: solar energy

Search Engine: gscholar

Year	Number of Hits
2000	7740
2001	8910
2002	9120
2003	10500
2004	11400
2005	12600
2006	14600
2007	16300
2008	18200
Total	

Figure 13: Hit Aggregator in Excel



INPUT SCREEN

Hit Aggregator

Search Term: Solar Energy

Beginning Year: 2004

End Year: 2008

Please Select Sources:

☑ Google Scholar

☑ Scirus

□ IEEE

□ Springer

□ Acm

□ Ingenta

□ Web of Science

☑ Plot Annual Chart

OUTPUT SCREEN

☑ Plot Annual Chart

Number of hits for search term: Solar Energy

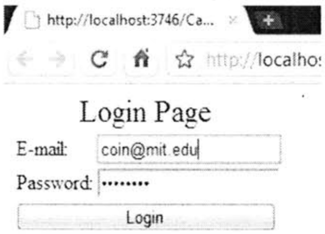
```
Google Scholar:11,400
Google Scholar:12,600
Google Scholar:14,600
Google Scholar:16,300
Google Scholar:18,100
Scirus:46,619
Scirus:55,074
```

Figure 14: Hit Aggregator Web Version

5.2 CAMELEON SCHEDULER

Cameleon Scheduler (CS) is a web application that allows users to register Cameleon queries and execute them according to a schedule specified through its web interface. The query results are recorded in a text file and can be downloaded by the user. CS can also be used to automate the execution of a parameterized query by supplying the parameters in a text file. Furthermore, CS offers automatic derivation of some parameter data. Details of these are explained below from the operational point of view.

CS can be accessed using a Web browser and the user first needs to login to the scheduler with an e-mail and password provided by the administrator (See Figure 15). The system will keep the user logged in for 30 minutes after a successful entry.



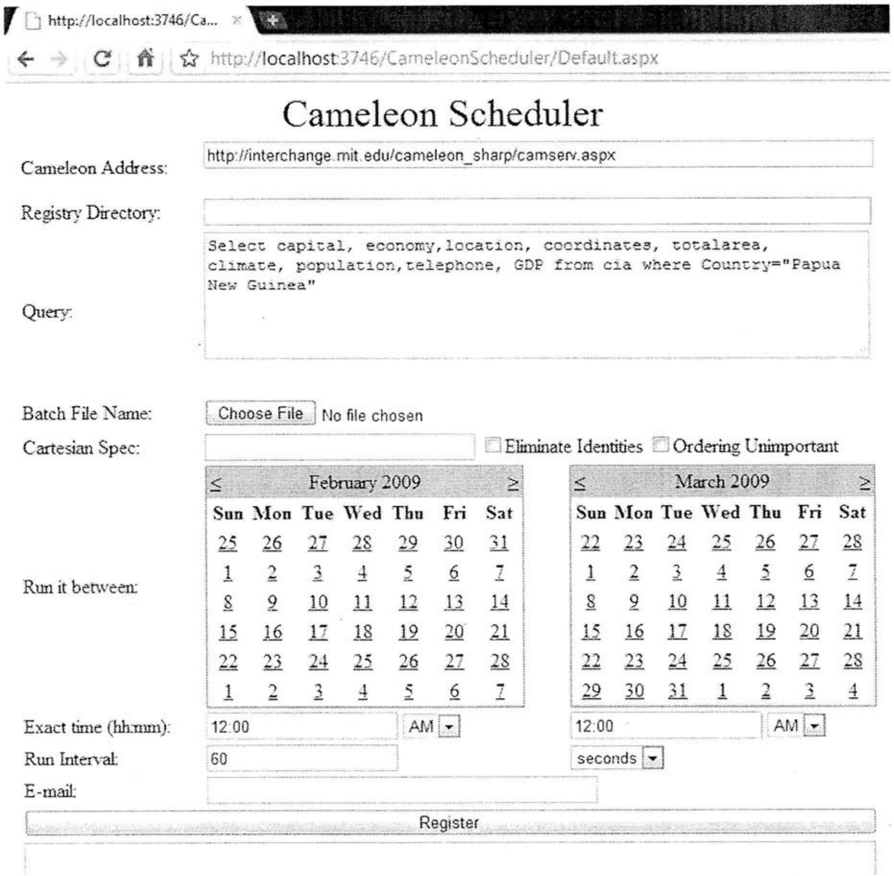
Login Page

E-mail: coin@mit.edu

Password: ••••••••

Figure 15: Login Page

Next, the user views the main CS web page. A snapshot of this page is shown in Figure 16 below. The first line in the figure is the address of the Cameleon engine the scheduler will utilize for execution of the query specified in the query box.



http://localhost:3746/CameleonScheduler/Default.aspx

Cameleon Scheduler

Cameleon Address: http://interchange.mit.edu/cameleon_sharp/camserv.aspx

Registry Directory:

Query:

```
Select capital, economy,location, coordinates, totalarea,
climate, population,telephone, GDP from cia where Country="Papua
New Guinea"
```

Batch File Name: Choose File No file chosen

Cartesian Spec: □ Eliminate Identities □ Ordering Unimportant

Run it between:

< February 2009 >

Sun	Mon	Tue	Wed	Thu	Fri	Sat
25	26	27	28	29	30	31
1	2	3	4	5	6	7
8	9	10	11	12	13	14
15	16	17	18	19	20	21
22	23	24	25	26	27	28
1	2	3	4	5	6	7

< March 2009 >

Sun	Mon	Tue	Wed	Thu	Fri	Sat
22	23	24	25	26	27	28
1	2	3	4	5	6	7
8	9	10	11	12	13	14
15	16	17	18	19	20	21
22	23	24	25	26	27	28
29	30	31	1	2	3	4

Exact time (hh:mm): 12:00 AM 12:00 AM

Run Interval: 60 seconds

E-mail:

Figure 16: Main Cameleon Scheduler Web Interface

In the registry box the user can enter a custom web accessible registry location in which the spec files referred in the query can be accessed. When left empty the registry points to the default registry location of the Cameleon engine specified in the first line.

In the query box the user specifies the query to be executed with a schedule. This query can be completely static: the same query is executed on specified intervals. Presumably, this could be desirable when the changing results of a query are to be tracked. Users can also express dynamic queries, which are parameterized and parameter values are obtained from a file. Consider for example the query:

```
Select capital, economy, location, coordinates, totalarea, climate,
population, telephone, GDP from cia where Country="$1"
```

and a countries.txt text file that contains the following entries:

Algeria
France
Germany
Poland
Turkey

The $1 parameter in the query refers to the first column of the countries.txt file, which needs to be specified as a batch file. In general $X refers to the Xth column of the batch file. Sometimes the user may supply a Cartesian product specification that transforms an existing file into another. If we specify $1x$1 as our Cartesian specification for the above countries.txt file, the file is transformed into a new one with two columns through Cartesian product as follows:

No box is checked	Eliminate identities box is checked	Ordering unimportant box is checked
Algeria, Algeria	~~Algeria, Algeria~~	~~Algeria, Algeria~~
Algeria, France	Algeria, France	Algeria, France
Algeria, Germany	Algeria, Germany	Algeria, Germany
Algeria, Poland	Algeria, Poland	Algeria, Poland
Algeria, Turkey	Algeria, Turkey	Algeria, Turkey
France, Algeria	France, Algeria	~~France, Algeria~~
France, France	~~France, France~~	~~France, France~~
France, Germany	France, Germany	France, Germany
France, Poland	France, Poland	France, Poland
France, Turkey	France, Turkey	France, Turkey
Germany, Algeria	Germany, Algeria	~~Germany, Algeria~~
Germany, France	Germany, France	~~Germany, France~~
Germany, Germany	~~Germany, Germany~~	~~Germany, Germany~~
Germany, Poland	Germany, Poland	Germany, Poland
Germany, Turkey	Germany, Turkey	Germany, Turkey
Poland, Algeria	Poland, Algeria	~~Poland, Algeria~~
Poland, France	Poland, France	~~Poland, France~~
Poland, Germany	Poland, Germany	~~Poland, Germany~~
Poland, Poland	~~Poland, Poland~~	~~Poland, Poland~~
Poland, Turkey	Poland, Turkey	Poland, Turkey
Turkey, Algeria	Turkey, Algeria	~~Turkey, Algeria~~
Turkey, France	Turkey, France	~~Turkey, France~~
Turkey, Germany	Turkey, Germany	~~Turkey, Germany~~
Turkey, Poland	Turkey, Poland	~~Turkey, Poland~~
Turkey, Turkey	~~Turkey, Turkey~~	~~Turkey, Turkey~~
Total: 25	Total:20	Total: 10

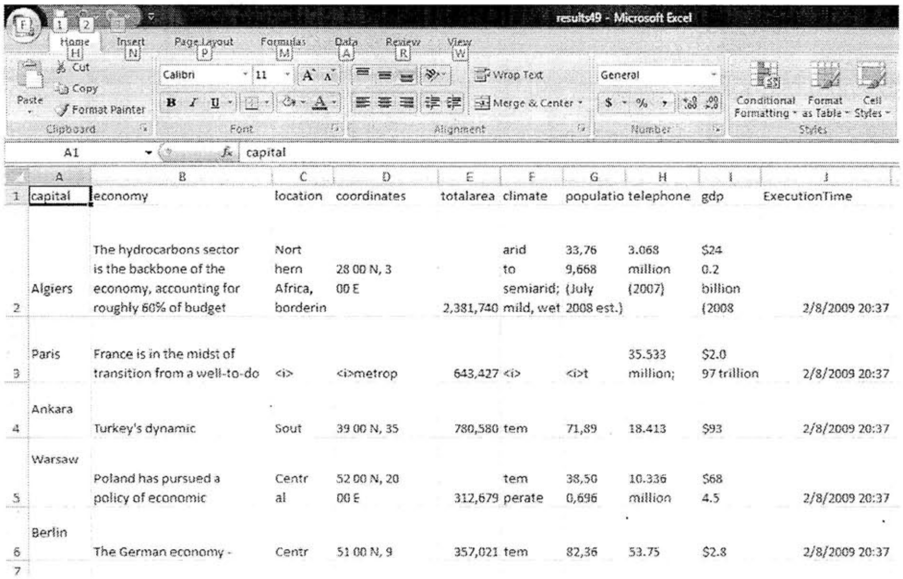
The user can then treat this transformed file as the loaded file and refer to its columns in the query. The Cartesian product is not limited to self product as in $1x$1 but can also take the forms such as $1x$2, $1x$3, $1x$2x$3, etc. assuming the referred columns exist in the loaded batch file. Note that columns in batch files are separated by commas.

Run it between line has two calendar selections that allow the user to specify the beginning and end dates of the schedule. These dates are coupled with the exact time of the execution listed just beneath them with default values of 12:00 AM.

The run interval will indicate how often the query (or queries as in the case of batch file specification) will be executed. This interval can be specified in seconds, minutes, hours, or days. If the run interval is greater than the difference between the beginning and end dates of the schedule, the query/queries are executed only once.

Note that when a batch file is specified, a list of queries is scheduled to execute periodically. If the intention of the user is to simply run a list of queries dynamically generated from a batch file once, the run interval should be a value that is larger than the difference between the beginning and end dates of the schedule.

The results of the execution are sent to the user as an attachment in an email in Excel (csv) format. For example, the results of the execution for the cia example are shown in Figure 17 below:



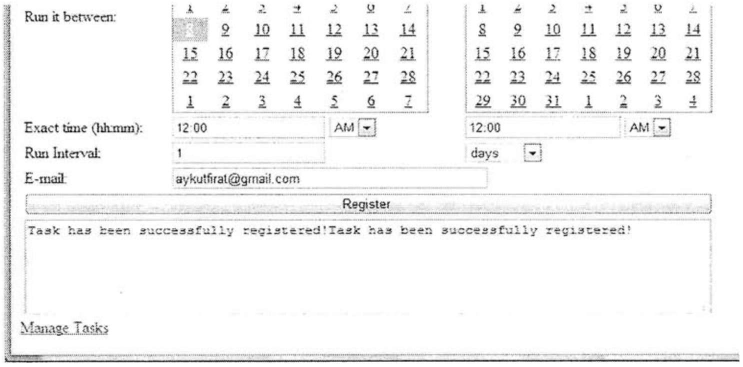
A1 capital

	A	B	C	D	E	F	G	H	I	J
1	capital	economy	location	coordinates	totalarea	climate	populatio	telephone	gdp	ExecutionTime
2	Algiers	The hydrocarbons sector is the backbone of the economy, accounting for roughly 60% of budget	Nort hern Africa, borderin	28 00 N, 3 00 E	2,381,740	arid to semiarid; mild, wet	33,76 9,668 (July 2008 est.)	3.068 million (2007)	$24 0.2 billion (2008	2/8/2009 20:37
3	Paris	France is in the midst of transition from a well-to-do	<i>	<i>metrop	643,427	<i>	<i>t	35.533 million;	$2.0 97 trillion	2/8/2009 20:37
4	Ankara	Turkey's dynamic	Sout	39 00 N, 35	780,580	tem	71,89	18.413	$93	2/8/2009 20:37
5	Warsaw	Poland has pursued a policy of economic	Centr al	52 00 N, 20 00 E	312,679	tem perate	38,50 0,696	10.336 million	$68 4.5	2/8/2009 20:37
6	Berlin	The German economy -	Centr	51 00 N, 9	357,021	tem	82,36	53.75	$2.8	2/8/2009 20:37
7										

Figure 17: Execution results for the CIA Factbook example

5.2.1 Managing Tasks

Currently, the system has a very simple task management system that allows users to delete existing tasks by clicking on the Manage Tasks link at the bottom of the interface (see Figure 18).


Run it between:

8	9	10	11	12	13	14
15	16	17	18	19	20	21
22	23	24	25	26	27	28
1	2	3	4	5	6	7

8	9	10	11	12	13	14
15	16	17	18	19	20	21
22	23	24	25	26	27	28
29	30	31	1	2	3	4

Exact time (hh:mm): 12:00 AM 12:00 AM

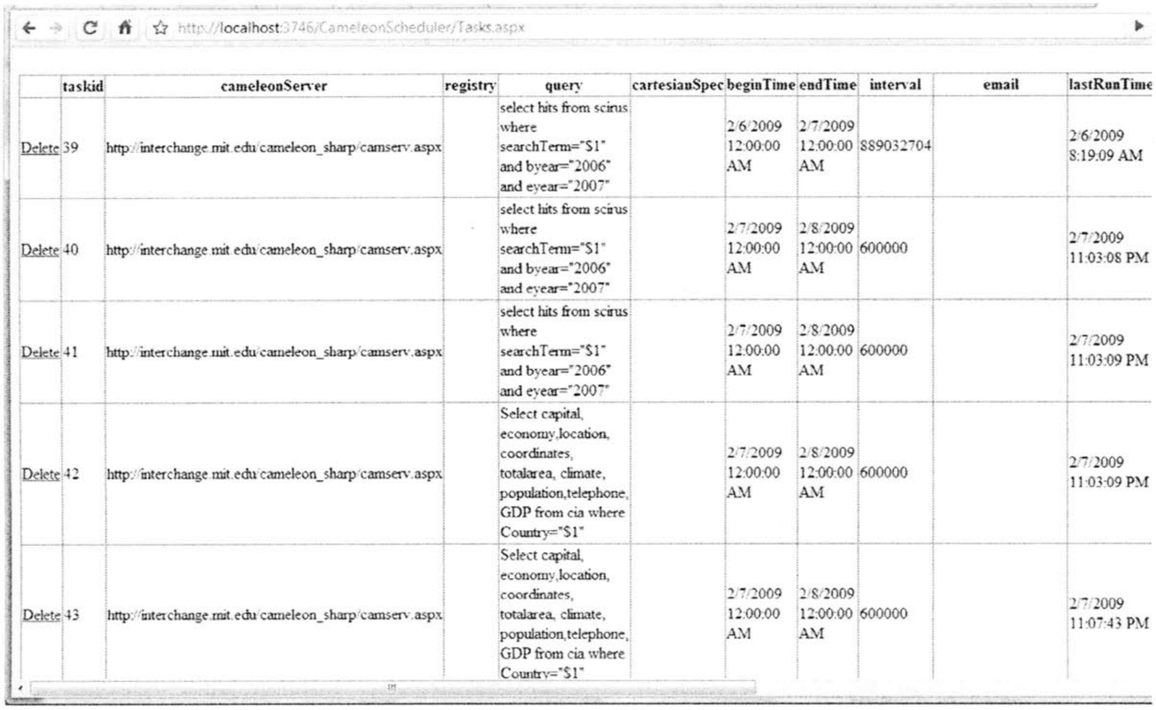
Run Interval: 1 days

E-mail: aykutfirat@gmail.com

```
Task has been successfully registered!Task has been successfully registered!
```

Figure 18: Manage Tasks Link

Task management allows users to view detailed information on already scheduled tasks and delete them. A sample snapshot is shown below in Figure 19.



http://localhost:3746/CameleonScheduler/Tasks.aspx

	taskid	cameleonServer	registry	query	cartesianSpec	beginTime	endTime	interval	email	lastRunTime
Delete	39	http://interchange.mit.edu/cameleon_sharp/camserv.aspx		select hits from scirus where searchTerm="$1" and byear="2006" and eyear="2007"		2/6/2009 12:00:00 AM	2/7/2009 12:00:00 AM	889032704		2/6/2009 8:19:09 AM
Delete	40	http://interchange.mit.edu/cameleon_sharp/camserv.aspx		select hits from scirus where searchTerm="$1" and byear="2006" and eyear="2007"		2/7/2009 12:00:00 AM	2/8/2009 12:00:00 AM	600000		2/7/2009 11:03:08 PM
Delete	41	http://interchange.mit.edu/cameleon_sharp/camserv.aspx		select hits from scirus where searchTerm="$1" and byear="2006" and eyear="2007"		2/7/2009 12:00:00 AM	2/8/2009 12:00:00 AM	600000		2/7/2009 11:03:09 PM
Delete	42	http://interchange.mit.edu/cameleon_sharp/camserv.aspx		Select capital, economy,location, coordinates, totalarea, climate, population,telephone, GDP from cia where Country="$1"		2/7/2009 12:00:00 AM	2/8/2009 12:00:00 AM	600000		2/7/2009 11:03:09 PM
Delete	43	http://interchange.mit.edu/cameleon_sharp/camserv.aspx		Select capital, economy,location, coordinates, totalarea, climate, population,telephone, GDP from cia where Country="$1"		2/7/2009 12:00:00 AM	2/8/2009 12:00:00 AM	600000		2/7/2009 11:07:43 PM

Figure 19: Task Management Interface

5.2.2 Technical Specifications

Cameleon Scheduler is a web application using ASP .NET. We used Access as the database to store task information mostly for ease of portability and lack of need for high concurrency. In the current version the data model in Access is shown in Figure 20 below.

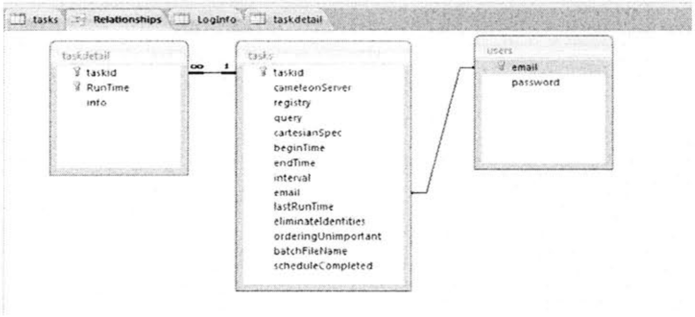


Figure 20: Data model for Cameleon Scheduler

The *tasks* table stores task information and *taskdetail* table stores information about each run (i.e. whether it was successful or failed). Users table will later be used to implement a more sophisticated user authentication system.

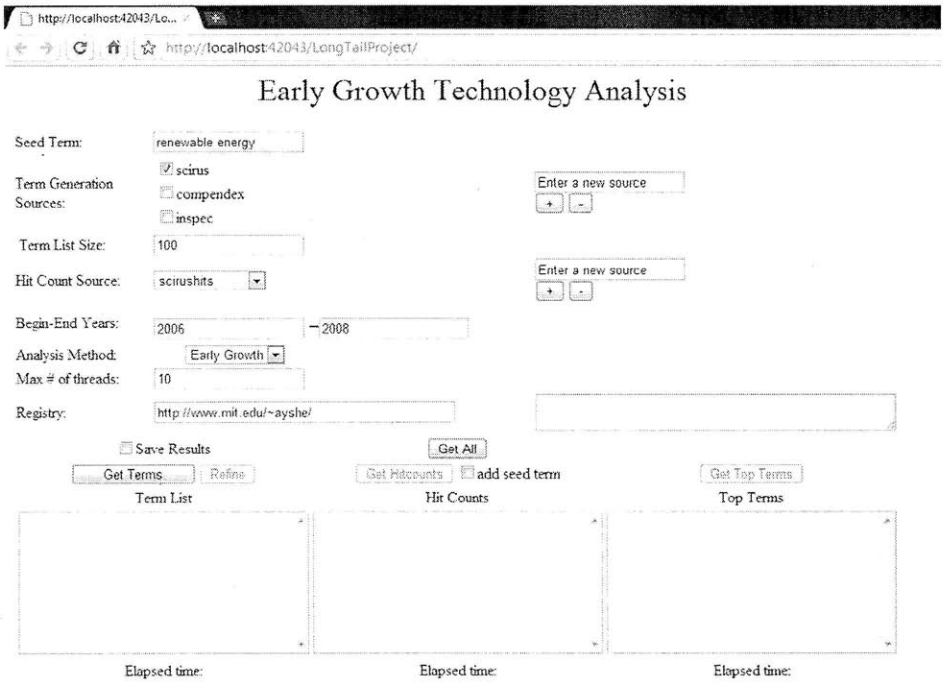
At a very high level, the scheduler operates as follows. When the user enters the necessary information and clicks on the register button, the data is recorded in the database. If a batch file was specified this file is uploaded to the server. If Cartesian specification was entered the file is transformed before being uploaded.

At the same time a timer event is running in the background that pools the data from the database every 30 seconds (this can be changed) and checks if there is any task that needs to be executed. If there is, the task is executed and its results are appended to its results file. If a batch file was specified, a list of queries is created first using the parameters and each of them are executed one by one.

Finally, when a task fulfills its lifetime, the results are sent to the user via email as an attachment. (See Appendix 3 to see how to set the email server in a configuration file.)

5.3 THE EGTA TOOL

The early growth technology analysis (EGTA) tool collects the required inputs from the user through a visual interface, extracts data from web sites on the fly, performs an analysis on the collected data, and displays the results. The user can interact with the tool either through a web-based or a desktop application interfaces (see Figures 21 & 22 below).



http://localhost:42043/LongTailProject/

Early Growth Technology Analysis

Seed Term: renewable energy

Term Generation Sources: ☑ scirus ☐ compendex ☐ inspec

Term List Size: 100

Hit Count Source: scirushits

Begin-End Years: 2006 – 2008

Analysis Method: Early Growth

Max # of threads: 10

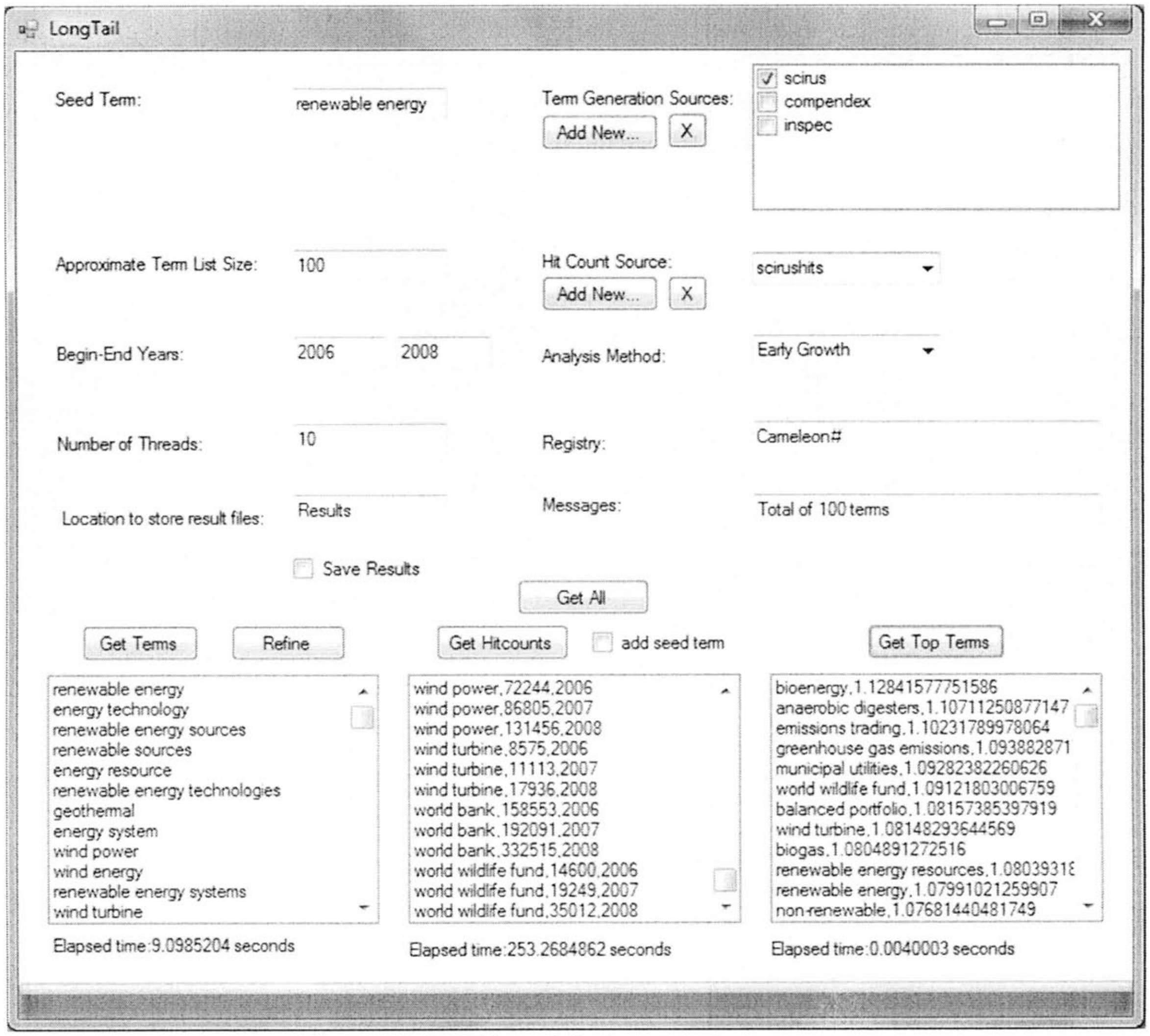
Registry: http://www.mit.edu/~ayshe/

☐ Save Results

☐ add seed term

Figure 21: EGTA tool web version

When the user enters a seed term, selects one or more term generation sources, a maximum term list size parameter, and hits the "Get Terms" button, related terms collected from the specified sources are shown in the term list box. When the user selects a single hit count source, specifies an interval in years, and hits the "Get Hitcounts" button, hit count of each term from the term list box and for each year that falls in the time interval are displayed inside the hit counts box. Finally, by hitting the "Get Top Terms" button, terms are listed in the top terms box in descending order of importance defined by the analysis method. Maximum number of threads limits outgoing parallel connections, and registry specifies where to get the spec files for external sources. The user can add new sources; save results; refine the results in the term box. Seed term can optionally be included in the queries sent to the hit count source, and all of the steps can be executed sequentially by clicking the Get all button.



Seed Term: renewable energy

Term Generation Sources: ☑ scirus ☐ compendex ☐ inspec

Approximate Term List Size: 100

Hit Count Source: scirushits

Begin-End Years: 2006 2008

Analysis Method: Early Growth

Number of Threads: 10

Registry: Cameleon#

Location to store result files: Results

Messages: Total of 100 terms

☐ Save Results

☐ add seed term

```
renewable energy
energy technology
renewable energy sources
renewable sources
energy resource
renewable energy technologies
geothermal
energy system
wind power
wind energy
renewable energy systems
wind turbine
```

Elapsed time:9.0985204 seconds

```
wind power,72244,2006
wind power,86805,2007
wind power,131456,2008
wind turbine,8575,2006
wind turbine,11113,2007
wind turbine,17936,2008
world bank,158553,2006
world bank,192091,2007
world bank,332515,2008
world wildlife fund,14600,2006
world wildlife fund,19249,2007
world wildlife fund,35012,2008
```

Elapsed time:253.2684862 seconds

```
bioenergy,1.12841577751586
anaerobic digesters,1.10711250877147
emissions trading,1.10231789978064
greenhouse gas emissions,1.093882871
municipal utilities,1.09282382260626
world wildlife fund,1.09121803006759
balanced portfolio,1.08157385397919
wind turbine,1.08148293644569
biogas,1.0804891272516
renewable energy resources,1.0803931ع
renewable energy,1.07991021259907
non-renewable,1.07681440481749
```

Elapsed time:0.0040003 seconds

Figure 22: EGTA tool desktop version

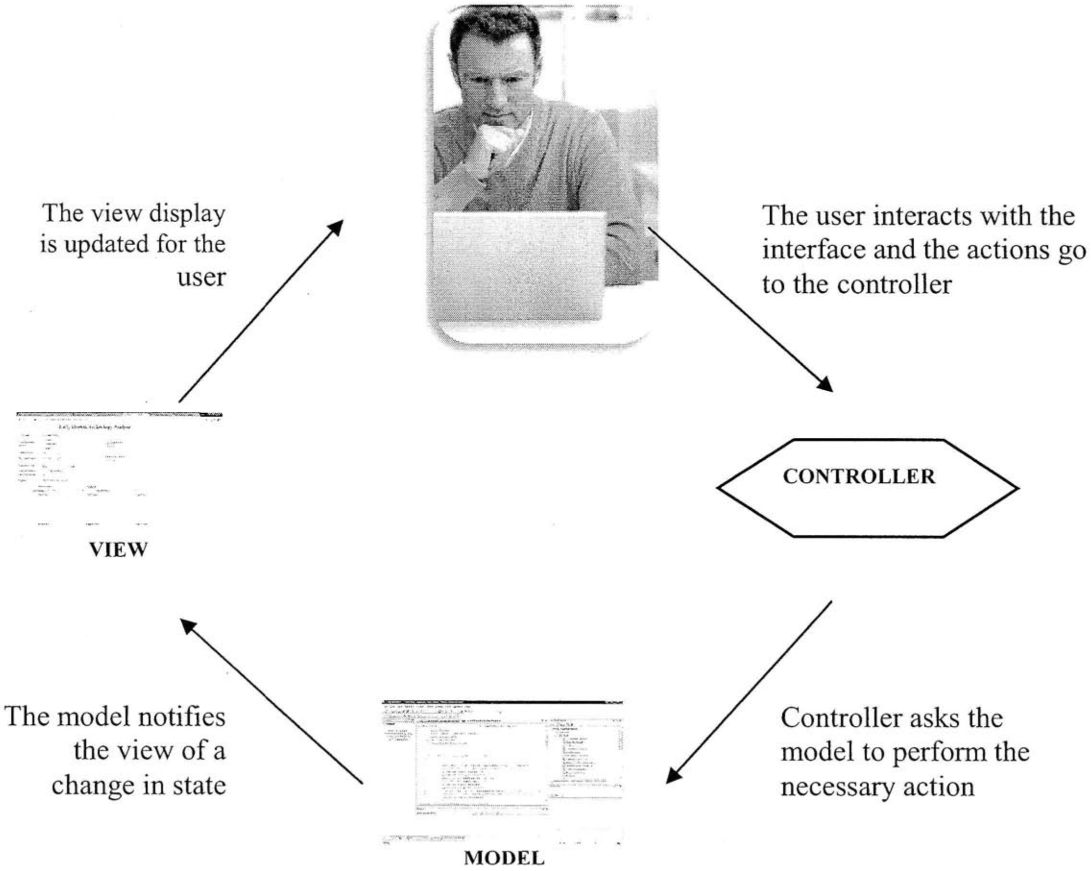
The desktop version is almost identical to the web version in terms of interface. A notable difference is that it includes a progress bar at the very bottom, displaying progress as the tool runs. Also the registry can be either a local or web directory, and the results are stored in a local directory.

Although the interfaces of the web and desktop versions are nearly identical, the applications have different uses. Through the web-based interface the user can access it from anywhere with an Internet connection. The desktop application, however, offers more control on the resources and can be run in a non-shared medium.

In the rest of this document we are going to describe the architecture of the system and provide installation and implementation details. The interested reader may find the code details in Appendices 4 and 5.

5.3.1 Software Design Details

We used the Model-View-Controller (MVC) design pattern[6] in developing the EGTA tool. MVC decouples the user interface (view) from the core code (model) by minimizing the dependencies between them using a controller. This process is illustrated in Figure 23 below. Using an MVC design pattern allows us to easily implement different interfaces without touching the model; thus preventing maintenance nightmares along the way.


The view display is updated for the user
The user interacts with the interface and the actions go to the controller
CONTROLLER
VIEW
The model notifies the view of a change in state
Controller asks the model to perform the necessary action
MODEL

The model contains all the state, data, and application logic needed to maintain and execute early growth analysis.

Figure 23: The MVC design pattern –
MVC allows us to exchange one view with another without touching the core

In both systems the model part of the code is identical and organized as shown in Figure 24. We used several sub design patterns within the MVC such as the observer pattern that allows subscribed listeners to be notified for errors and progress.

[6] In software engineering, a design pattern is a general reusable solution to a commonly occurring problem in software design. (Wikipedia)

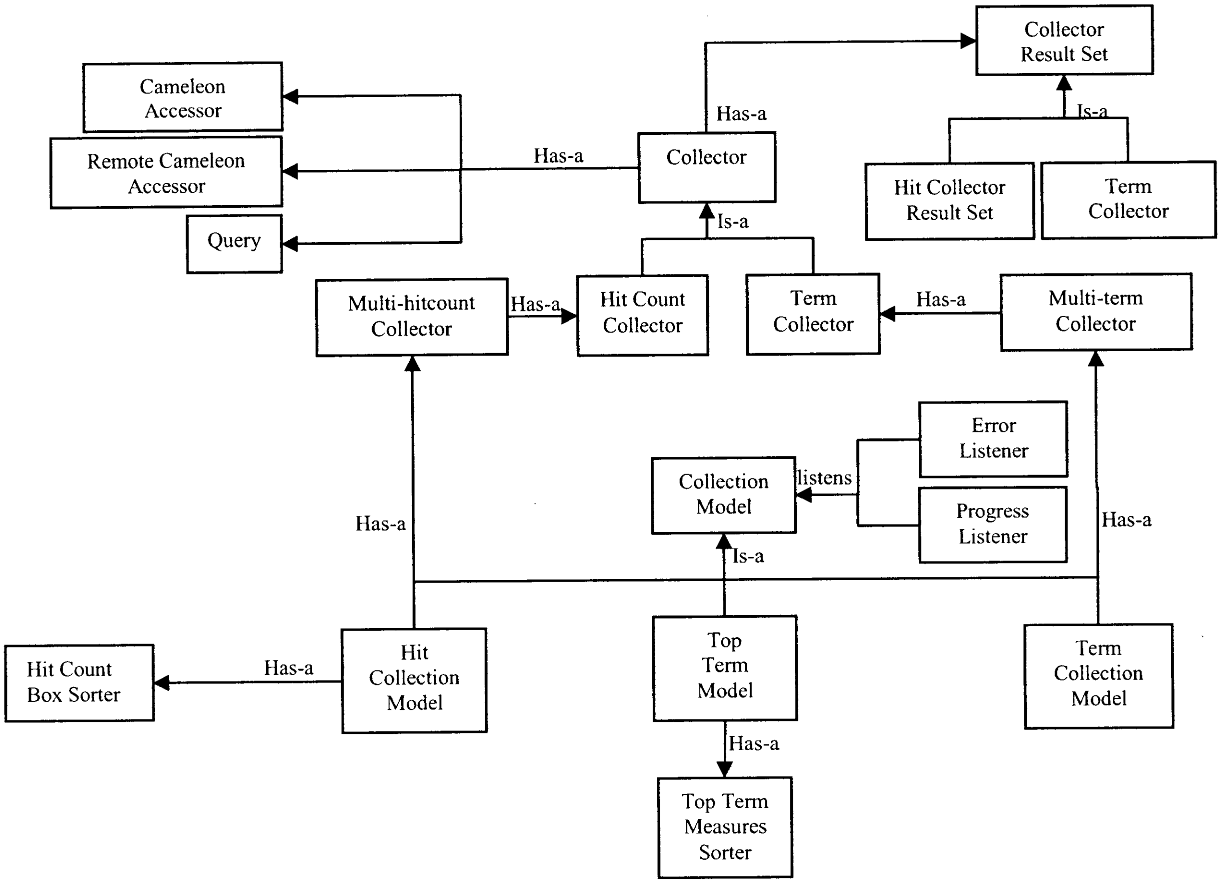


Figure 24: The Model Classes

Both the Controller and the View consists of one class(not shown in Figure 24), which are different for the desktop and web applications.

5.3.2 Software Implementation

We have implemented both applications in C# under the freely available Visual Studio Express (C# and Web) integrated development environments (IDE). The system runs in ASP .Net 3.5 framework. The applications also run under Unix environments with the help of Mono project. The complete code can be found in Appendix 5. We will discuss below some of the implementation issues.

Local vs Remote Cameleon Access

Cameleon is a web extraction tool we utilize in our applications. Currently, we use a local Cameleon (as a dll) to access remote sources, except in the cases of sources that can only be accessed through MIT. For those sources our applications use the Cameleon server that is in Interchange.mit.edu.

Concurrency in Web Source Access

Originally we were using both the POE (parallel version of Cameleon) and Cameleon for access to the web sources. Because of some technical issues related to POE and the desire to hide implementation details from the user we currently only use Cameleon with a default thread size of 10.

Formulas for Top Term Calculation

We currently use the following formulas for the calculation of top term ranks.

Early Growth = Log(hits in End Year) / Log(hits in Beginning Year)

Total Growth = (hits in End Year - hits in Beginning Year)/hits in Beginning Year)

Progress Bar in Web Application

Although we have a progress bar in the desktop application, we do not have an equivalent one in the web application. Because of the difficulty of asynchronous web programming, using a progress bar is not very straightforward, thus left out in the current implementation.

Adding New Sources

New sources can be added by uploading spec files to the registry directory. This can be done locally in the desktop application, or remotely for the desktop application. The specFiles.txt and specFilesHit.txt files (found under the registry directory in the desktop application, and the root folder in the web application) keep track of existing spec files in the system. When the user adds new sources through the interfaces these files are updated. Note, however, that in the desktop application we rely on the user putting the appropriate files in the registry directory after doing the add operation.

CHAPTER 6: SOLAR ENERGY CASE STUDY

Case studies using our tool enable us showcase the promise of our approach. It is also a chance to evaluate the relevancy of results, and offer clues about further improvement. Case study results also offer a chance to compare the sources used in the analysis. By comparing the intersection of term lists, for example, we can judge whether the sources are complementary or substitutes of each other. This chapter presents the results of applying Early Growth Technology Analysis (EGTA) to the field of solar energy.

We present two term lists related to the solar energy. We manually and painstakingly created the first list by sifting through many online data sources[7], journal articles, and taking the opinions of experts on solar energy. Our aim was to find a list of solar technologies that are currently novel and emerging. We also perform back-testing by examining the hit counts of these technologies in the last decade. Our expectation is to catch an early growth pattern for the technologies that are currently hot.

The second list is automatically generated by our EGTA tool. We compare this auto-generated list to the manually created list to understand how much overlap there is. Although, the auto generated list my point to technologies that are yet to be hot, we still expect to find some level of overlap.

We also compare our term-generation data sources (Scirus, Inspec and Compendex), by analyzing how the generated term lists correlate. Finally, we discuss the emerging research topics that are uncovered by our software tool.

6.1 MANUALLY CREATED EMERGING SOLAR TECHNOLOGIES LIST

Since we did not have expertise in the solar energy field, finding emerging and novel solar technologies was an arduous process that required countless hours. Identifying technologies that are not yet commercialized and still in the lab is not an easy task for the novice of the field. After sifting through online sources and expert opinions, we came up with the following solar technology list as the hot solar technologies of today:

1. Copper-indium-gallium di-selenide (CIGS) pv cells
2. Quantum wells
3. Quantum wires
4. Quantum dot solar cells
5. Hot carrier solar cells
6. Formation of intermediate band gaps
7. Flexible high-performance silver metallization
8. Amorphous silicon thin film solar cells

[7] Some of the sources used are:
"A Strategic Research Agenda for Photovoltaic Solar Energy Technology." Working Group 3 "Science, Technology and Applications" of the EU PV Technology Platform, 2007
"Technology Roadmap Solar photovoltaic energy." The International Energy Agency (IEA), 2010
"Special Report: Solar Power." MIT Technology Review, 2010
"EU PVSEC: Fraunhofer ISE develops new solar PV technologies with greater efficiency". RenewableEnergyFocus.com News, September 2009
"Flexible Thin-Film Technology: A Novel Metallization Paste." RenewableEnergyWorld.com Article, March 2009

9. Cadmium telluride (CdTe) thin film cells
10. Cadmium telluride (CdTe) with silicon thin film cells
11. n-type silicon solar photovoltaics (PV)
12. Advanced inorganic thin films
    - Spherical copper indium di-selenide (CIS) approach
    - Polysilicon thin film (polysilicon Si solar cell)
13. Organic solar cells / organic PV
    a. Hybrid approaches in which organic solar cells retain an inorganic component (e.g. the Graetzel cell)
        - Dye solar cells
        - Nanocrystalline dye-sensitised solar cell modules
    b. Full-organic approaches
        - Bulk donor-acceptor heterojunction solar cells,
        - Full-organic bulk heterojunction solar cells made with screen-printed transparent contact and active layer (requires improved and stable polymers, stabilization of nanomorphology, organic multi-junctions)
14. Thermophotovoltaics (TPV)
15. Flexible array of light-absorbing silicon microwires and light-reflecting metal nanoparticles
16. Solar Air Conditioning
17. Direct wafer making

6.2 BACKTESTING

An important question to ask about our manually constructed hot solar technology list of the previous section is as follows: "Do these technologies display a growth behavior that can be caught with a hit count analysis?" To find an answer to this question, we collected hit counts for most of the technologies in our list by using our hitcount aggregator and Google Scholar as our hit count provider. It is interesting to note that this type of hitcount analysis describes the prevalence of well-known technologies such as RFID remarkably well. For example, the hitcount curve of RFID as shown below in Figure 25 indicates that RFID had limited prevalence prior to 2001, then underwent "early growth" from roughly 2001-2003, rapid growth between 2003-2006, and has been relatively flat afterward.

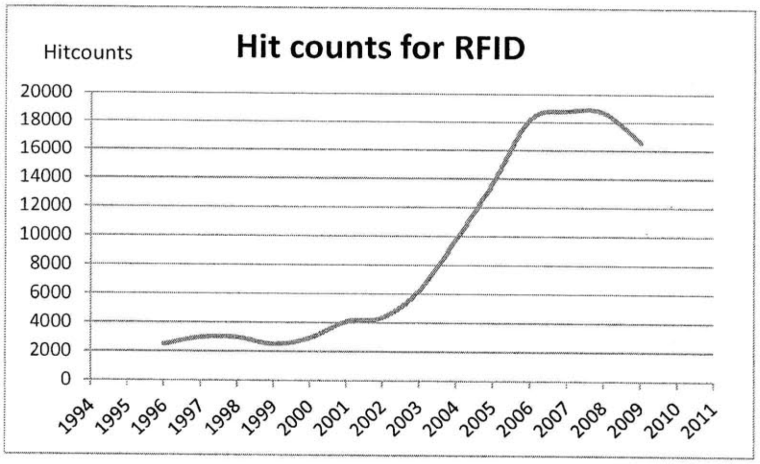


Figure 25: RFID Hitcount Growth between 1996-2009

The hit count charts for 12 of the technologies listed in the previous section in the last 10 years are shown in Figure 25. Note that x-axis represents year, and y-axis corresponds to hitcounts in the graphs.

When we examine these graphs, we easily find the early growth pattern in 7 of the 12 cases ("Quantum dot solar cells", "Copper Indium Gallium di Selenide", "Hot carrier solar cells", "Amorphous Thin Silicon Solar Cells", "Cadmium Telluride", "Silver Metallization", "Dye solar cells"). In 4 of the 12 cases, we see flattish behavior ("Quantum wells", "Quantum wires", "N-type Silicon", "Poly-silicon thin film". The case of "thermophotovoltaics" is inconclusive because there are relatively very few (<100) hitcounts over the years. This result answers the question in the opening mostly affirmatively, and signals that our assumptions about early growth do indeed have the potential to uncover emerging technologies through hit count analysis.

6.3 COMPARISON OF SOURCES

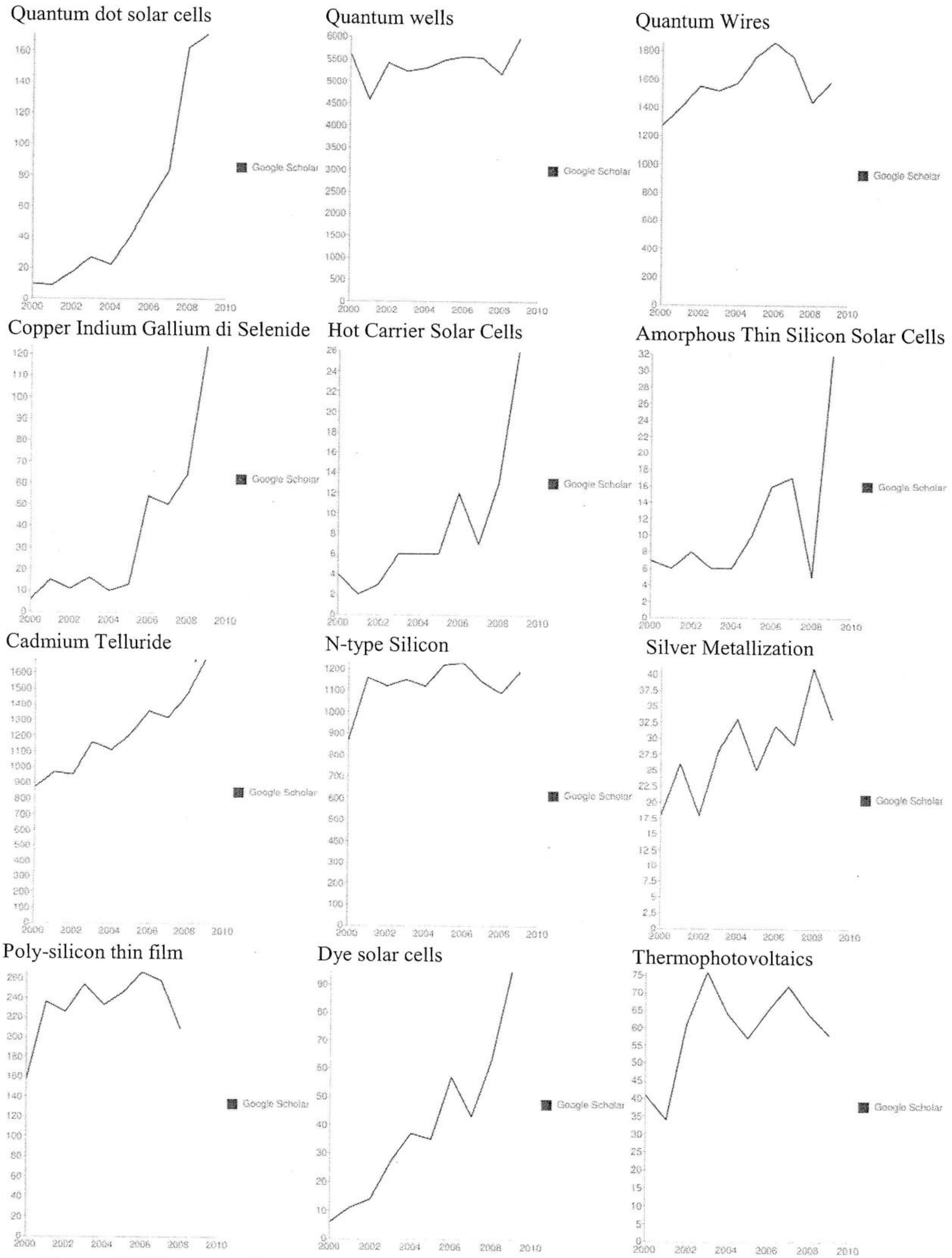
We used three different online journal databases, Scirus, Compendex and Inspec, to collect the solar energy related terms for the case study presented in this chapter. In this section, we analyze the similarity of the term lists generated by these sources. Understanding similarities and differences between online databases of scientific publications may be important in calibrating future studies.

To perform this analysis, we used seed terms with different target term-list sizes and calculated the overlap percentage of the results in Compendex, Scirus and Inspec. As shown in Table 7 below, the overlap percentages are very low for these three resources, ranging from 4% to 26%. This might mean that these resources index very different types of information, and they can be treated as complementary sources.

Seed Term – Term Size	Compendex/Inspec	Compendex/Scirus	Inspec/Scirus
Photovoltaics -100	11%	4%	8%
Photovoltaics - 1000	18.6%	7.9%	6.9%
Photovoltaics -10000	25%	6.3%	5%
Solar energy - 500	16.4%	6%	4.4%
Solar energy - 3000	26.1%	6.5%	4.2%
Renewable Energy - 100	13%	4%	4%

Table 7: Overlap percentage of results for seed terms "Photovoltaics", "Solar energy", and "Renewable energy" with different term sizes between Compendex, Scirus and Inspec.

In a secondary analysis, we also found out differences in the way data are organized in these databases. We chose seed terms with similar meanings (such as gasoline and petrol, PV and Photovolatcis) and generated term lists for each of the seed terms. Then we calculated the intersection of the related term lists of each seed term. Higher intersection is more desirable because we expect similar terms to return approximately similar results. We found out that term lists generated by Scirus with related terms had less intersection than Compendex and Inspec as shown in Table 8. From this brief analysis, it is clear that differences in the collection methods as well as in the underlying data sources can result in very different keyword sets.


Quantum dot solar cells

Quantum wells

Quantum Wires

Copper Indium Gallium di Selenide

Hot Carrier Solar Cells

Amorphous Thin Silicon Solar Cells

Cadmium Telluride

N-type Silicon

Silver Metallization

Poly-silicon thin film

Dye solar cells

Thermophotovoltaics

Figure 26: Hit counts for twelve emerging technologies from Google Scholar for the last ten years.

Source	Solar Energy and Photovoltaics	PV and Photovoltaics	Gasoline and Petrol
Compendex 2100	74%	54%	64%
Inspec 2100	65%	48%	75%
Scirus 2100	30%	34%	33%

Table 8: Overlap percentage of results from Compendex, Inspec and Scirus for term size 2100 for pairs with similar or same meaning (e.g. Gasoline and Petrol)

6.4 CASE STUDY RESULTS

In this section, we compare the manually created emerging solar technologies term list of section 6.1 with the term list we obtained automatically from our EGTA tool. As reported below, there are important overlaps between the two lists, which is an encouraging sign for our research. There is, however, also room for improvement especially in getting rid of irrelevant terms from the list, finding better ranking schemes, and combining relevant terms together using hierarchies.

Below we report some of our observations:

6.4.1 Metals

The literature review on emerging and novel solar technologies reveal that metals like Gallium, Selenium and Copper are quite important for the field. In our manually created list we had:

Copper-indium-gallium di-selenide (CIGS) pv cells
Cadmium telluride (CdTe) thin film cells
Cadmium telluride (CdTe) with silicon thin film cells
Spheral copper indium diselenide (CIS) approach

Our EGTA tool lists these metals several times:

Rank	Term
228	Selenium
450	Cadmium Compounds
814	Gallium
884	Copper Compounds
893	Gallium Alloys
909	Gallium Arsenide
929	Semiconducting Selenium Compounds
946	Semiconducting Gallium
1011	Gallium Nitride
1186	Cadmium Alloys
1405	Indium Sulfide
1416	Indium
1432	Semiconducting Indium

6.4.2 Nanotechnology

Improvements in the field of nanotechnology are quite important for solar energy. Emerging technologies of section 6.1 list *nanostructures* and *nanoparticles*. Ray Kurzweil's reported talk in Chapter 2 also emphasized the intersection of nanotechnology and solar energy. Our EGTA tool lists nanotechnology related terms very often and with high ranks:

Rank	Term
8	Nanophotonics
20	Nanosheets
39	Nanotechnology
66	Nanocomposites
125	Nanostructured Materials
156	Nanowires
172	Nanocrystals
264	Nanorods
362	Nanoparticles
368	Nanostructures

6.4.3 Quantum dots, wires and wells

Quantum dots, *wires* and *wells* that are mentioned in our manually constructed emerging technologies list come up in our EGTA results:

Rank	Term
186	Quantum Chemistry
199	Quantum Optics
377	Semiconductor Quantum Dots
818	Quantum Well Lasers
950	Semiconductor Quantum Wells
955	Quantum Theory
1005	Quantum Well Infrared Photodetectors
1292	Quantum Efficiency

6.4.4 Formation of intermediate band gap

Terms related to "formation of intermediate band gap"s in our emerging technologies list can also be tracked in our results:

Rank	Term
798	Ultra-Wideband (Uwb)
846	Photonic Band Gap
1958	Optical Band Gaps

6.4.5 Thin films

Terms similar to the emerging technologies 'Amorphous silicon thin film solar cells', 'Polysilicon thin films', 'Full-organic bulk heterojunction solar cells', 'Direct wafer making', 'Hot carrier cells' appear in our list.

Rank	Term
14	Organometallics
38	Semiconducting Intermetallics
133	Photorefractive Crystals
576	Heterojunctions
1119	Wafer Bonding
1156	Organic Polymers
1175	Semiconducting Organic Compounds
1399	Semiconductor Junctions
1414	Silicon Wafers
1544	Amorphous Silicon
1738	Polysilicon
1776	Microcrystalline Silicon
1821	Carrier Concentration

6.4.6 General solar energy related terms

Our results also have general solar energy related terms such as:

Rank	Term
132	Solar Water Heaters
197	Solar Power Generation
206	Silicon Solar Cells
207	Solar Energy
212	Solar Concentrators
220	Mirrors
236	Solar Power Plants
274	Solar System
312	Solar Collectors
318	Solar Equipment
380	Ultrathin Films
399	Solar Refrigeration
658	Solar Heating
677	Solar Buildings
705	Solar Radiation
949	Solar Absorbers
1059	Solar Cell Arrays
1157	Thin Films
1586	Passive Solar Buildings
1653	Thin Film Transistors
1659	Concentration (Process)

6.5 CONCLUSION

In this chapter we compared a manually created term list corresponding to emerging solar technologies with the term list auto generated by our EGTA tool. We found out that there is a great degree of overlap between these lists, which is an encouraging sign for our research. We also performed a limited back testing, which indicated that in most cases we can expect to find the early growth pattern by examining the hitcounts of an emerging technology over the years. Furthermore, we identified that the sources we used return largely different result term lists for the same seeed term; therefore can be treated as complementary sources. Compendex and Inspec also seem to handle synonyms more effectively compared to Scirus. Understanding the differenced and similarities of data sources may be important in designing future studies. Complete term list returned by our EGTA tool can be viewed in Appendix 6.

CHAPTER 7: GEOTHERMAL CASE STUDY

In this chapter, we present the results of applying Early Growth Technology Analysis (EGTA) to the field of geothermal energy. In Appendix 7, we present 2000 terms found and ranked by the EGTA tool for the seed term "geothermal". We have larger data to report (for 4000 terms, and for 5000 terms), but by eyeballing the results, it seems that after 2000 terms, the list starts including more generic terms and less interesting technologies.

The term list was created using Compendex and the hit counts were created by Scirus. We selected Compendex for term generation, because the term list included more specific technologies than generic terms when compared to the term lists generated by Scirus. For hit counts, however, we preferred Scirus, because it has a wider collection of scientific publications. We categorize the results of geothermal case study into 4 main areas and discuss them below. These four categories indicate whether the automatically generated terms from the EGTA tool are related to the geothermal field, related but generic terms, unrelated generic terms, or totally unrelated terms. Such a classification helps us understand the quality of the results returned by the EGTA tool.

7.1 CATEGORIZATION OF THE CASE STUDY RESULTS

A. Technologies that might be related to the geothermal field

These results are promising in that they include many relevant terms we would hope to find on such a list based on our brief analysis of the field as reported in Section 7.2.

Rank	Term
180	Osmosis
460	Ammonia
488	Supercritical Fluids
511	High Performance Liquid Chromatography
727	Solvent Extraction
774	Water Cooling systems
1128	Protective Coatings
1180	Flow simulation
1251	Ion Exchange Resins
1374	Enhanced Recovery
1385	Multi/Single-Walled Carbon Nanotubes (Swcn)[8]
1508	Large Eddy Simulation[9]
1509	Thermal Logging
1531	Hydraulic Fracturing[10],

[8] ***Carbon nanotubes (CNTs)*** are allotropes of carbon with a cylindrical nanostructure. Nanotubes have been constructed with length-to-diameter ratio of up to 28,000,000:1 which is significantly larger than any other material. These cylindrical carbon molecules have novel properties that make them potentially useful in many applications in nanotechnology, electronics, optics and other fields of materials science, as well as potential uses in architectural fields.

[9] ***Large eddy simulation*** **(LES)** is a numerical technique used to solve the partial differential equations governing turbulent fluid flow (Wikipedia)

[10] ***Hydraulic fracturing*** is a method used to create fractures that extend from a borehole into rock formations, which are typically maintained by a proppant, a material such as grains of sand or other material which prevent the fractures from closing. The method is informally called fracking or hydro-fracking.

1761	Plasma Enhanced Chemical Vapor Deposition
1848	Evaporative Cooling Systems
1991	Fins (Heat Exchange)

B. Generic terms related to the geothermal field:

We also find terms whose level of generality makes them undesirable.

Rank	Term
111	Fossil Fuels
112	Earth (planet)
140	Earthquakes
174	Turbines
1715	Volcanic rock

While these terms are all somehow relevant to geothermal, the connection may be loose or apply equally well to unrelated seed terms or be far from specific technologies that are in their early growth stages.

C. Technologies unrelated to the geothermal field

Our list also includes specific terms that are unrelated to the geothermal field:

Rank	Term
175	Wireless Sensor Networks
472	Particle Accelerators
728	Cellular Telephone Systems
783	Computer Operating Systems

D. Generic terms unrelated to the geothermal field

We also observe generic terms that are unrelated to the geothermal field:

Rank	Term
68	Surgery
72	Education
77	Housing
91	Linguistics
102	Insulin
168	Metabolism
181	Antibiotics
943	Body Fluids

We then looked at the top 20 terms returned by our EGTA tool, and assigned one of the four categories we described above. This helps us understand the relevancy of our EGTA results. Note that we do not have expertise in geothermal energy, and our classification is a based on a best-effort understanding of the domain using online sources.

Rank	Terms	Category
1	Myoelectrically Controlled Prosthetics	C
2	Refractometers	B
3	Magnetrons	A

4	Fragrances	D
5	Phototransistors	C
6	Nanofibers	A
7	Nanofluidics[11]	A
8	Quartz Crystal Microbalances[12]	A
9	Abrasives	A
10	Aneroid Altimeters	B
11	Electron Microscopes	C
12	Navigation	B
13	Microscopes	D
14	Observatories	D
15	Plastic Molds	B
16	Steel Metallography	A
17	Tsunamis	B
18	Adsorbents	A
19	Tanning	B
20	Navigation Systems	B

Although the list includes some items that are not directly related to what we are looking for, it is encouraging to find 7 highly relevant terms in the top 20 list. For a more comprehensive examination of the results, we present in the next section, a brief survey on promising geothermal technologies, which can be compared to the complete term list returned by the EGTA tool and presented in Appendix 7.

[11] ***Nanofluidics*** is the study of the behavior, manipulation, and control of fluids that are confined to structures of nanometer (typically 1-100 nm) characteristic dimensions (1 nm = 10-9 m). Fluids confined in these structures exhibit physical behaviors not observed in larger structures, such as those of micrometer dimensions and above, because the characteristic physical scaling lengths of the fluid, (e.g. Debye length, hydrodynamic radius) very closely coincide with the dimensions of the nanostructure itself.

[12] ***Quartz Crystal Microbalances*** A quartz crystal microbalance (QCM) measures a mass per unit area by measuring the change in frequency of a quartz crystal resonator. The resonance is disturbed by the addition or removal of a small mass due to oxide growth/decay or film deposition at the surface of the acoustic resonator. The QCM can be used under vacuum, in gas phase ("gas sensor", first use described by King[1]) and more recently in liquid environments. It is useful for monitoring the rate of deposition in thin film deposition systems under vacuum.

7.2 BRIEF ANALYSIS OF THE EMERGING TECHNOLOGIES IN GEOTHERMAL ENERGY

We can classify the promising new developments in geothermal surface technology[13] as follows:

I. Incremental technology improvements that marginally increase the efficiency of a particular power plant component, for example—might prove most valuable in the near term. Geothermal power plants can always benefit from reduced parasitic load, reduced power expenditures related to cooling fans, improvements to the power substation, and other modest technological advances.

II. Increasingly Standardized, Modular Geothermal Conversion Systems: ***Modularity*** allows developers to more easily add capacity after a reservoir has been found to be capable of additional production.

Related terms in our automatically generated list were:

Rank	Term
1273	Phase Modulation
1803	Multicarrier Modulation
1875	Pulse Width Modulation

III. Mineral Recovery: Further research and development could make the separation of minerals from geothermal water, known as mineral recovery, a viable technology. Some geothermal fluids contain significant concentrations of dissolved minerals, while others are virtually mineral free. Mineral recovery offers several benefits, which generally fall into categories of either improving the function of the power plant (reducing scaling, allowing greater power production by lowering the injection temperature), or increasing profits (through the sale of mineral byproducts). Often a variety of benefits will result. Minerals found at geothermal power plants include zinc, silica, lithium, manganese, boron, lead, silver, antimony and strontium.
(eg. a combination of already existing technologies modified for the task: ion exchange, ***solvent extraction***, and "electrowinning" to extract ***zinc*** from the used geothermal liquid.)
(eg. Preliminary results suggested in 2006 that ***silica*** recovery at Mammoth Lakes could reduce the cost of geothermal electricity production by 1.0¢/kWh. The market value of silica that could be produced from the Mammoth Lakes site if silica is removed from all geothermal liquid is estimated to total $11,000,000/year.) LLNL is also considering using ***reverse osmosis*** to separate lithium, cesium, rubidium and tungsten. However, these activities have not yet been pursued.

Related terms in our automatically generated list were:

Rank	Term
66	Biomineralization
127	Minerals
162	Mineralogy
180	Osmosis

[13] The State of Geothermal Technology, Part II: Surface Technology, Geothermal Energy Association for the U.S. Department of Energy, 2008.

253	Organic Minerals
587	Zinc
599	Silicate Minerals
711	Oxide Minerals
727	Solvent Extraction
1028	Mineral Oils
1140	Silica
1158	Zinc Sulfide
1213	Zinc Oxide
1347	Sulfate Minerals
1458	Organic Solvents
1459	Semiconducting Zinc Compounds
1466	Fused Silica
1520	Phosphate Minerals
1978	Tantalate Minerals

IV. Working Fluids for Rankine Cycle Power Plants: Studies at National Renewable Energy Laboratory (NREL), Idaho National Laboratory(INL), and elsewhere have shown that mixed working fluids in binary-cycle geothermal power plants can potentially reduce thermodynamic inefficiencies in the boiler and condenser, thereby improving overall plant efficiency. Researchers have investigated various pure and mixed working fluids to optimize power conversion efficiency. One potential working fluid, especially suitable for lower temperature resources, is ***ammonia*** water as used in the Kalina cycle.

Related terms in our automatically generated list were:

Rank	Term
23	Nanofluidics
72	Microfluidics
150	Superfluid Helium
460	Ammonia
488	Supercritical Fluids
585	Computational Fluid Dynamics
781	Fluidized Beds
1093	Cascades (Fluid Mechanics)
1698	Fluid Structure Interaction

V. Hybrid Cooling Systems: Experts cite hybrid ***cooling*** as one of the most important areas for surface improvements. Due to the increasing demand for both efficiency and water resources, NREL and INL investigated ways to improve the heat transfer effectiveness of air-cooled condensers. The NREL concept, according to the lab's R&D website, "involves the use of perforated fins in which all air flows through the perforations. Tests of two prototypes at NREL and associated computer modeling indicated that 30 to 40 percent more heat transfer could be obtained for the same fan power with a hybrid as opposed to a stand-alone air cooling system."

Related terms in our automatically generated list were:

Rank	Term
661	Cooling Systems
1848	Evaporative Cooling Systems

VI. Coatings: Corrosion and deposition of mineral scale (known as fouling) can occur at geothermal resource areas with high concentrations of dissolved and suspended solids. When scale accumulates over time, it can clog pipes or vessels and decrease the effectiveness of heat exchangers.

In order to reduce the cost of maintaining open flow paths and efficient heat transfer, Brookhaven National Laboratory (BNL) developed durable, scale-resistant ***polyphenyl sulfide***-based coatings for carbon steel. These coatings can be used for heat exchangers—devices which transfer heat through a conducting wall from one fluid to another—as well as for binary cycle power plants, piping, flash vessels, and other plant components.

Related terms in our automatically generated list were:

Rank	Term
954	Phenolic Resins
1265	Thiophene
1611	Phenols
1877	Sulfide Minerals

7.3 CONCLUSION

Like the solar case study, the results for the geothermal study are very promising. We find many highly relevant terms in the top results returned by our EGTA tool, and few generic, and unrelated terms. Having generic and unrelated terms, unfortunately, cannot be avoided entirely when using an automated process; however there have only been relatively few occurrences of such terms. Also, the relevance of some terms may not be immediately obvious, but there might actually be some relationship. Haptic interfaces, for example, are widely used in virtual environments and tele-operator applications, and these could probably be quite useful for geothermal exploration.

Having implemented the EGTA approach and completed two case studies we talked to some local experts to get their reaction to our work. This is presented next, in Chapter 8.

CHAPTER 8: EXPERT INTERVIEWS

To better understand the utility of the work we described in this Thesis, we conducted interviews with three MIT affiliated experts in the technology forecasting area: William K. Aulet, Howard Anderson, and Satwik Seshasai. William K. Aulet, the acting managing director of MIT's entrepreneurship center, is a serial entrepreneur who has raised over $100 million in funding for his companies and directly created hundreds of millions of dollars of market value. Aulet believes that execution and implementation is more important than the selection of a particular technology, but he was quite excited about our work. He considers the ability to map technology landscape through bibliometric analysis as an important decision aid in the technology selection process.

Howard Anderson, the founder of The Yankee Group and the Co-Founder of Battery Venture Capital, is a Senior Lecturer at the MIT Entrepreneurship Center. He sits on several high technology boards in the communications, computing, and advanced materials industries, and recently was voted one of the top 25 people in the communications industry by Network World magazine. In our interview, Anderson stressed the importance of pattern recognition. He believes that VCs and scientists all use pattern recognition to make decisions, and the ability to pick up patterns with less and less data and use them as early predictors is of paramount importance to technology forecasting. He considers the work we have done as a good starting point that can attract considerable attention if we can show some positive correlation through back-testing.

Satwik Seshasai is a Senior Manager at IBM and also a PhD Candidate at Massachusetts Institute of Technology focusing on technology forecasting. Seshasai offers a different perspective on the utility of our work and other similar efforts. He believes that they are more useful on identifying how to craft a market message around the technology, on identifying how to reorganize existing groups to deliver the technology, and on deciding which companies to acquire as opposed to deciding what to do. For example, when building a medical product, this type of work can be useful in identifying complementary products; hence influence acquisition decisions. We learn more about interesting topics, the terminology of those topics, and other topics related to those topics.

Below we provide partial transcripts of two of the interviews to provide a more detailed account of their thoughts on technology forecasting in general, and our work in particular.

8.1 INTERVIEW TRANSCRIPTS

HOWARD ANDERSON - is the founder of The Yankee Group and the Co-Founder of Battery Venture Capital. He is a Senior Lecturer and MIT Entrepreneurship Center. At MIT, he teaches 15.390 New Enterprises, 15.398 Companies at the Crossroads, 15.386 Managing in Adversity and 15.387 High Technology Sales and Sales Management. He sits on several high technology boards in the communications, computing, and advanced materials industries. He recently was voted on of the top 25 people in the communications industry by Network World. His commentary can often be read in such publications as The Wall Street Journal, Forbes Magazine and Technology Investor.

[We start with a brief explanation of our research]

Howard Anderson: So your assumption is that the buzz for new technology shows up early in emails, searches, databases and that is a good predictor.

Ayse Kaya Firat: Our hypothesis is that prevalence might be a good indicator. We not only look at hit counts (prevalence) in online journal databases but also the growth rates. We look at historical trends in search for what the "emerging technologies" of the past looked like and try to learn and extrapolate from them. We are after "obscure" terms, the technologies even experts in a field may not be aware of.

HA: Sometimes the hype as a backlash and the actuality might be different. Let's take something like RFID: [HA draws a figure like the following]

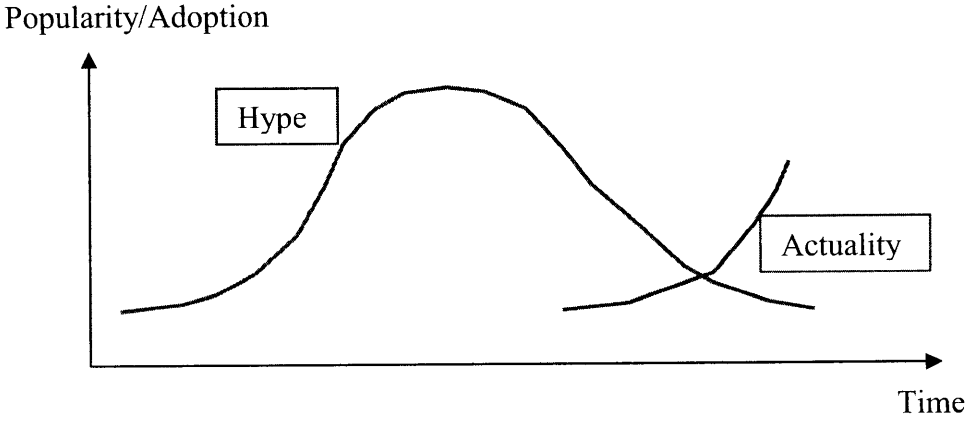


Figure 27: RFID Popularity/Adoption vs. Time Curve

As the hype levels of a technology goes down and reality goes up, people start saying "This is not really working." The transition from hype to reality might take 10 years. 5-6 years ago everyone was saying RFID was a great new technology, then they turned around and said "Hey, it is very expensive, and it does not work". Be careful when you see something and think it is a good leading indicator. I see this sometimes in standards and in communication where everyone is supposedly behind this and then nothing happens.

AKF: Do you think sources might make a difference? We look at online journal databases, academic articles. Not blogs or newspapers.

HA: What happens often is the academic community comes first, industry the second. Something will come out of chemical engineering department and we, venture capitalists will eventually catch onto it, and now we start to talk about it and once we made an investment into a company, we hype it ourselves. If someone calls me up from a magazine asking what the three new great emerging technologies are, I'll say "Lithium ion batteries", because I just put 2 million dollars in it. So that kind of gets along a life of its own. It starts with technical articles, then finds its way into the business articles, then all of a sudden it is on the cover of Business Week.

AKF: In your own experience, how is technology forecasting being done in companies?

HA: I ran a group called Yankee Group, and Yankee Group was in that business. I started the company in 1970 and sold it in 2000. By the time I left, I had 200,000 clients who would pay as much as $50,000 a year to get our view of which new companies and technologies were important. One of my company's job was to tell users about new technology inventors they should pay attention to, and tell the vendors what smart users were doing. So, kind of an information switch.

I was a popularizer. Why is a company paying this much amount of money to do this? Should not they be doing it themselves? Let's say a company like Goldman Sachs which has an IT budget that's almost $2 Billion. They say they want to know about a new technology that can be a differentiator, where do we learn that from? Do we read the journal articles ourselves? Probably not. But we pay other people to be forward scouts: Yankee Group, Gartner Group and Forrester being some of them. They influence other users and publications that write an article like "The Best 15 New Technologies".

AKF: How do you come up with such a list? Do you use any quantitative methods?

HA: Absolutely not. One area that I look is where I think there is going to be some unique companies. I look where I am putting venture capital, and where I think the market is going to be very good for some time, and what I think is going to be commercially successful. It's like a play. There is a lot of data running back and forth. Sometimes things will catch fire, sometimes there is a whole new platform, like the Internet. And sometimes there is false information that people think it's true: e.g. 'internet is doubling every year". More and more money goes in, then there is a vicious cycle.

AKF: But let's say IBM is thinking of investing in 10 technologies and there are millions of technologies out there. How do they go from a million to the top 100? I understand that once they have the top 100, they might use many methods, like cost-benefit analysis to go to top 10.

HA: I don't think there are millions. There are basic categories and subclasses. I do not know how exactly it is done, but IBM, for example makes a billion dollars a year in licensing fees. How do they do it? What they do is they say to all their divisions "You keep track of the competition and whatever licensing revenue we get goes to your profit center." What they do is, they rely on the engineers and scientists' in the divisions to keep track of what is going on in the market.

AKF: IBM has a team called HorizonWatch composed of about 1600 people who discuss the newest technologies on internal blogs, conference calls, etc.

HA: President of AT&T is a friend of mine, he called up and said "IBM is making a billion dollars, how do I do the same thing? I've got all these patents, too". I said "You need an organization to do this. Not full-time, that's not a full-time job." He just said "I don't have the time and the effort to do it". So there is a semi-formal way the people keep track.

Sometimes a company will say "Alright, there is this new technology, we have some corporate money to sponsor trial for 6 months. After that one of the divisions has to adopt it. If nobody wants it, move on to something else. So what you have is a kind of a trial and error method.

AKF: Do you think that these technology investment decisions are based on repeatable, objective criteria? How robust are they?

HA: Most of the innovations do not come from big companies, they come from little companies. I divide the world into two groups: Attackers and defenders. The defenders are the big technology companies. Let's say there is a new technology that comes up. Let's call it VoIP. The big companies probably get a few people playing with it a few years, but it's not doing too well; mainly because, it threatens the financial viability of their existence. They keep trying on it, but they do not push it. Little company pushes it, and they get some attraction. Now, the big company's first inclination is denial. The second inclination is anger. And the third thing is "Maybe we should oem the product, and put our label on it" and then, in the end they wind up selling it through an investment banker to a big company.

If you look at a company like Cisco, you'll see that they have done hundreds of acquisitions. Microsoft does the same thing. Corporate development guys always keep their eye on new technologies. This is what Boston Scientific is trying to do. They say "We're not smart enough to know which of these dozens of technologies are important. What we will do is we'll pick maybe 10 or 20 things that we think have high potential. We'll put in $5 million in each of these technologies.

AKF: So, experts come together and decide…

HA: Yeah, I've got a CTO, or a CMO who says this is kind of interesting. There is some risk. Let's put a few little bets out here. Maybe we'll have an agreement that we have the right to buy out 50% at three times sales in five years. So they say let's not smother the poor company. Let's give them some money, and if they are successful, we'll pay them a lot of money but not so much that it becomes unaffordable. And that sometimes works. The idea that I want to plant a lot of seeds and a few of them will grow. And no one is smart enough to know which particular seed is going to work. And sometimes they are dead wrong; they may make five bad investments.

If you look at companies like Pfizer or Merck, they spend 10 billion dollars a year on R&D for 10 years, and never have any great new products out there. Why? They just chose wrong, they were too risk averse, they bet on the wrong horse. They would do anything including buying your master thesis to find if there is a better way of predicting here. And maybe looking at blogs would be very worthwhile. They try bunches of things, but had not been successful.

Do you know how penguins find out if it is safe to go into water?

AKF: They push each other...

HA: Right, they push each other until somebody falls in and if the shark does not eat them, they all jump in. That's kind of what happens here. We exchange the same information, we go listen to Robert Langer speak. His investments count five times anybody else's. He is someday going to have more patents than Thomas Edison. We'll look through the patent applications and then we'll say this is the hot area. Then the investment money comes. If it turns out to be worthwhile, now there are 20 companies doing the same thing.

AKF: I talked to William K. Aulet of MIT's Entrepreneurship Center and, although he was quite excited about this project, he said technology is a little bit overrated. He said what really matters, especially in the energy sector is execution and implementation.

HA: Everyone says the same words. When you look at the energy sector, you may see a curve. Everyone gets really excited about certain things, and then it tails off. He is right in his way.

AKF: But let's say you carry out the same successful execution/implementation steps for two different scenarios. For the first one you start with a better portfolio of technologies than the second. Don't you think it will make a difference?

HA: You never have the same execution; the costs are different, the timing is different. I had a great idea for a new turbine engine. But now the government has price supports: Let's take Arnold Schwarzenegger says that if you drive an electric car, you can get in the high-speed lane in California. And all of a sudden the battery guys are doing a lot better. Now, the technologies that were marginally not worth doing, become more worth doing.

AKF: One last question: Do you think automated approaches like ours might be interesting to people out there? Are they even aware of approaches like bibliometric analysis and technology mining?

HA: I think the answer is it would be of interest if you can show some positive correlation. If not, it's a good first start. Everyone is smart enough to know that they better pay attention to staff that comes out of a university. It may not be perfect the first time, but maybe it gets better.

We're all in the business of pattern recognition. I wrote an article about why VCs and scientists both use pattern recognition to make decisions. Can we pick up patterns with less and less data? And you are looking if there is an early predictor. And it might be a trailing indicator. You'll look at that and see if you can use your data to do a simulation. You'll go back to 2003 and see what you would have predicted and then you look at reality. Was that a decent predictor? And if it works out, then we'll pay money for you to do more research.

AKF: Thanks a lot, Prof Anderson.

SATWIK SESHASAI - Senior Manager at IBM and PhD Candidate at Massachusetts Institute of Technology

Ayse Kaya Firat: My first question is about how technological forecasting is done in practice.

Satwik Seshasai: Many companies use terms like 'Market analysis', 'Market Intelligence' and 'Insight' instead of calling it technology forecasting. The market intelligence groups are very disconnected from the product groups. They produce reports and suggestions, and the product groups build products independently. I think this is general in the industry. What market intelligence does almost by accident can influence product groups.

AKF: Who does the market intelligence group report to?

SS: That's the thing right. They report separately. By the way, reporting structure does not mean anything in those companies. Let's say Polycom who is producing telephones may have an engineering team, which is building the next version of the desktop phone. And then they have a market analysis group that is looking at devices in general, the strategy: should it be small, should it be big, should it be mobile, so forth. By the time they influence the product group, the product group may have already made certain decisions. You asked how decisions are made. It is not like everybody sits down in a room and say "Alright. Now we're going to make a decision." There is never actually a decision point, things just happen. The strategy group might say "We think you should make these devices wireless with 3G enabled, so you do not need to hook it up to a wire". For that to get from there to the product group, it is kind of by accident. They produce a report and say this is where the market is going. The engineers may (or may not) decide to look at the report and say "Let's experiment with that". In parallel some customer may hear about it and says "I really want it", and request a prototype. As the engineers are already building the prototype, the prototype goes to the customer, and if they like it, the prototype enters the product line. But it is not like there was a decision point.

AKF: How do they justify the existence of a market intelligence group, then?

SS: By the size. A small company would not do this. But a large company says 'we should spend X% of our budget on this, because it is generally good to have.

AKF: IBM has a group called Horizon Watch. Can you explain what Horizon does?

Basically, they are a group inside IBM's market intelligence that identifies specific areas to look into such as intelligent transportation. These areas are not aligned with the product groups. They are more looking outwards. They are saying what is happening in the industry. Based on what is happening in the industry, let's go explore. They do not care if IBM already has a product in the area or not.

AKF: How do they identify the areas?

SS: By feeling. Once they identify the areas, they say who is interested among the whole company. People who are interested or who has expertise, participate in discussion groups, in online forums and conference calls. It is a very ad hoc aggregation of data. Someone might say "I read this, have you seen this article?" But there is no rigorous methodology behind how they operate. At the end, all they produce is some aggregation and analysis of the information in that domain.

AKF: Are you aware of any systematic studies like bibliometric analysis, publication citation analysis, etc?

SS: There may be individuals within the company who choose that as a methodology to provide their own points. I may choose to browse the web. And we both speak up at the conference call.

AKF: What about IBM's market intelligence? Do they use quantitative methods?

SS: They do, but it is usually reusing what analyst firms like Gartner or IDC do. These analyst firms look at spend, interview customers and different vendors, and so forth.

If they look at which technologies are better than the other, that's all expert opinion with very tactical quantitative inputs. Should we invest in this chemical or that chemical? There will be people who say this is better or that is better. Then, there may be some study that says the market for products based on that chemical is this big. There is nothing like 'because this has grown in citation versus that, we're going to invest in that instead of that'.

AKF: Some say selection of technology is overrated and what matters is implementation and execution. How do you think your own thesis and our work in general will contribute?

SS: I believe all of this work is not useful for deciding on the technology, it's more useful on identifying how to craft a market message around the technology, on identifying how to reorganize existing groups to deliver the technology, on deciding which companies to acquire and so forth. All of those types of decisions are where this type of work is useful as opposed to deciding what to do. Let's say you are building this medical product, is there a complementary product that's built by a company that you need to acquire? Such questions can be answered by this. What this says is more people are talking about this topic versus that topic, they are using this term as opposed to this term, when they talk about this they also talk about that.

AKF: Thanks, Satwik.

CHAPTER 9: CONCLUSION

In this Thesis, we presented Early Growth Technology Analysis (EGTA), an automated bibliometrics based method to aid decision makers in identifying novel technologies that have the potential to generate high commercial returns at an early stage.

9.1 CONTRIBUTIONS

Our main contribution is the creation and implementation of a decision aid tool that takes advantage of the knowledge buried in online scientific publications to help decision makers and experts identify emerging technologies in a field they are interested in; ensure that they do not miss an important development; augment and validate their already established ideas on options to consider for investment and fund allocation.

Given the expectation that this decision aid tool may be used by executives at our sponsoring organizations, IBM and Masdar, ease of maintenance and extendibility of the software was a primary goal. We accomplished this goal by adopting design patterns in software design, which make further development and maintenance much easier. Although there is still a lot of room for improvement, we augmented the decision aid tool with visual interfaces to improve the total user experience.

We also advanced the state of the data collection within our group by separating the generic and source-specific aspects of data collection, which are no longer buried together in some black-box program; and are cleanly separated. Ability to declare source-specific data collection knowledge without coding improves the extendibility and maintenance of the EGTA software tools.

The two case-studies we performed using the EGTA indicate that our methodology is promising in identifying technologies that are in their early growth phase. Emerging technologies identified by EGTA mostly contained the hot technologies we found via an arduous process of sifting through many relevant publications, and expert opinions in the solar and geothermal energy fields. We got encouraging responses from the experts we interviewed with on the applicability of our approach as a decision aid tool in technology forecasting.

9.2 FUTURE WORK

The work presented in this Thesis can be followed up with additional work in several directions. First, a thorough study can be done to compare different formulas to measure early growth. In our study, we relied on a formula that measured the growth using a base and a final year. Deriving the shape of the growth curve by including intermediate years, for example, may improve the ranking of the resulting technologies even though it comes with a performance penalty because of increased data collection.

Similarly, the quality of term collection can be improved by finding an optimum combination of depth and breadth parameters; including more sources for term collection; and by utilizing the knowledge about the coverage of sources for different domains.

An important area being investigated by researchers in our group is automatically constructing taxonomies for the collected terms. A related effort can be undertaken on how to use taxonomies to improve the EGTA. Taxonomies may improve the ranking schemes, for example, by aggregating the hit counts of terms that are synonyms or closely related, or part of a term group. They can also be utilized to guide the term collection process. Although we did not elaborate in this Thesis, we extended the EGTA tool to interact with the automatic taxonomy creation algorithms, but we did not experiment with it extensively.

Finally, we view using alternative sources as another avenue to build upon our research. In addition to online databases of scientific publications, alternative sources such as patent databases, and faster changing online data sources such as blogs can be used to improve the performance of EGTA.

With further enhancements, we believe that EGTA will be a promising data-driven approach to be used by policy makers to complement, verify, and validate expert opinions in coming up with funding allocation and technology investment decisions.

REFERENCES

[Baker 2008] Baker, J. "Masdar Tools Report," Internal Memorandum, MIT Sloan School of Management, April 2008.

[Brown and Phaal 2001] Brown R., Phaal, R. The use of technology roadmaps as a tool to manage technology developments and maximize the value of research activity, IMechE Mail Technology Conference (MTC), Brighton, 2001.

[Camiña 2010] Camiña, S., A Comparison of Taxonomy Generation Techniques Using Bibliometric Methods: Applied to Research Strategy Formulation. Composite Information Systems Laboratory (CISL), Massachusetts Institute of Technology, 2008.

[Campbell 1966] Campbell, R., A methodological study of the utilization of experts in business forecasting, PhD Dissertation, UCLA, 1966.

[Coates et al. 2001] Coates, V., Faroque, M., Klavins, R., Lapid, K., Linstone, H.A., Pistorius, C., and Porter, A.L. On the future of technological forecasting, Technol. Forecast. Soc. Change 67 (1) 1 – 17, 2001.

[De Miranda Santo et al. 2006] De Miranda Santo, M., Coelhno, G. M., Dos Santos, D.M., and Filho, L. F. Text mining as a valuable tool in foresight exercises: a study on nanotechnology, Technol. Forecast. Soc. Change 73, pp. 1013–1027, 2006.

[Dalkey 1969] Dalkey, N. An experimental study of group opinion: The Delphi method, Futures 1 408 – 420, 1969.

[Dalky et al. 1963] Dalky, N., Helmer, O. An experimental application of the Delphi method to the use of experts, Manage. Sci. 9 458 – 467.

[European Commission 2006] European Commission, Emerging Science and Technology Priorities in Public Research Policies in EU, US and Japan, pp 13-24 and 25-35, 2006.

[Farquhar 1970] Farquhar, J.A., A Preliminary Inquiry into the Software Estimation Process, Rand, RM-6271-PR, 1970.

[Firat 2008] Firat, A. Cameleon#1 User Manual. Composite Information Systems Laboratory (CISL), Massachusetts Institute of Technology, 2008.

[Firat et al. 2005] Firat, A., Madnick, S., Yahaya, N.A., Kuan, C.W, and Bressan,S. "Information Aggregation using the Caméléon# Web Wrapper," Proceedings of the 6th International Conference on Electronic Commerce and Web Technologies (EC-Web 2005)," Copenhagen, Denmark, August 23 - August 26, 2005, also published in Springer Lecture Notes in Computer Science (LNCS) 3590, K. Bauknecht et al (Eds.), pp.76-86, Springer-Verlag Berlin, 2005.

[Gordon 2003] Gordon, T.J. "A simple agent model of an epidemic," Technological Forecasting & Social Change,vol. 70 pp. 397-417. 2003.

[Gordon and Glenn 2003] Gordon, T.J., Glenn J.C. (Eds.), Futures research methodology, Version 2.0 Millennium Project of the American Council for the United Nations University, 2003.

[Gustafson et al. 1973] Gustafson, D.H., Shukla, R.K., Delbecq, A., Walster, G.W. A comparative study of differences in subjective likelihood estimates made by individuals, interacting groups, Delphi groups, and nominal groups, Organ. Behav. Hum. Perform. 9, 280 – 291. 1973.

[Hill and Fowless 1975] Hill, K.Q., and Fowles, J. The methodological worth of Delphi forecasting technique, Technol. Forecast. Soc. Change 7, 179 – 192. 1975.

[Hood and Wilson 2001] Hood W., and Wilson, C.S. The Literature of Bibliometrics, Scientometrics, and Informetrics. Scientometrics 52(2):291–314, 2001.

[Levary and Han 1995] Levary, R., Han, D. Choosing a technological forecasting method. Industrial Management, 37(1), 14, 1995.

[Lichtenthaler 2004a] Lichtenthaler, E. Technological change and the technology intelligence process: a case study, J. Eng. Technol. Manag. 21 (4), pp. 331–348, 2004.

[Lichtenthaler 2004b] Lichtenthaler, E. Technology Intelligence Processes in Leading European and North American Multinationals. R&D Management 34 (2) 121–135, 2004.

[Martino 1970] Martino, J. The precision of Delphi estimates, Technol. Forecast. 1, 293 – 299. 1970

[Martino 1983] Martino, J. Technological Forecasting for Decision Making, 2nd ed. New York, NY: North-Holland, pp: 39, 40. 1983.

[Millett 2003] Millett, S. The future of scenarios: challenges and opportunities, Strategy and Leadership 31 (2)16–24, 2003.

[Mishra et al. 2002] Mishra, S. Deshmukh, S.G., Vrat, P., "Matching technological forecasting technique to a technology", Technol. Forecast. Soc. Change (TFSC) 69, pp. 1–27, 2002.

[National Science Foundation 2010] National Science Foundation, Proposal Review and Processing, http://www.nsf.gov/bfa/dias/policy/meritreview/phase2.jsp 2010.

[Norling et al. 2000] Norling P.M., Herring J.P., Rosenkrans, W.A., Stellpflug M, and Kaufman S.B., "Putting Competitive Technology Intelligence To Work" Research-Technology Management, Vol. 43, No. 5, pp. 23-28, 2000.

[Parente et al. 1984] Parente, F., Anderson, J., Myers, P., and O'Brien, T. An examination of factors contributing to Delphi accuracy, J. Forecast. 3 173 – 182, 1984.

[Pfeiffer 1968] Pfeiffer, J. New Look at Education: System Analysis in Our Schools and Colleges, Odyssey, New York, 1968.

[Phillips et al. 2005] Phillips, J.G., Heidrick, T.R., Potter, I. J. "Technology Futures Analysis Methodologies for Sustainable Energy Technologies", Technology Management: A Unifying Discipline for Melting the Boundaries 155-165. 2005.

[Porter 1998] Porter, A.L. Innovation forecasting using bibliometrics, Compet. Intell. Rev. 9 (4), pp. 11–19, 1998.

[Porter 2007] Porter, A.L. P. in van der Duin (ed.), Knowing Tomorrow? How Science Deals with the Future, Eburon Academic Publishers, Delft 183-201. 2007.

[Porter and Detampel 1995] Porter, A.L. and Detampel, M.J. Technology opportunities analysis, Technol. Forecast. Soc. Change 49, pp. 237–255. 1995.

[Porter and Cunningham 2005] Porter, A.L., and Cunningham, S.W. Tech Mining: Exploiting New Technologies for Competitive Advantage, John Wiley & Sons, Hoboken (New Jersey) 2005.

[Rabinow 2004] Rabinow, P. "Assembling Ethics in an Ecology of Ignorance" Presented as the closing plenary lecture of the First Conference on Synthetic Biology, MIT, 10-12 June 2004.

[Reger 2001] Reger, G. "Technology foresight in companies: from an indicator to a network and process perspective." Technology Analysis & Strategic Management, 13, 4, 533– 553, 2001.

[Sahlo and Cuhls 2003] Sahlo, A., and Cuhls, K. "Technology foresight – past and future", Journal of Forecasting, Vol. 22. pp.79-82. 2003.

[TFAMWG 2004] Technology Futures Analysis Methods Working Group (TFAMWG), Technology futures analysis: toward integration of the field and new methods, Technol. Forecast. Soc. Change 71 (3), pp. 287–303. 2004.

[Woon and Madnick 2008(1)] Woon, W., Madnick. S. Asymmetric Information Distances for Automated Taxonomy Creation. Composite Information Systems Laboratory (CISL), Massachusetts Institute of Technology, 2008.

[Woon and Madnick 2008(2)] Woon, W., Madnick. S. Technology Forecasting Using Data Mining and Semantics, MIT/MIST Collaborative Research, 2008-2009.

[Woon et al 2009(1)] Woon, W., Henschel, A., Madnick, S. A Framework for Technology Forecasting and Visualization. Composite Information Systems Laboratory (CISL), Massachusetts Institute of Technology, 2009.

[Woon et al. 2009(2)] Woon W., Zeineldin, H., Madnick, S. Bibliometric Analysis of Distributed Generations. Composite Information Systems Laboratory (CISL), Massachusetts Institute of Technology, 2009.

[Woundenberg 1991] Woundenberg, F. An evaluation of Delphi, Technol. Forecast. Soc. Change 40 131-150. 1991.

[Ziegler 2009] Ziegler, B. Methods for Bibliometric Analysis of Research: Renewable Energy Case Study. Composite Information Systems Laboratory (CISL), Massachusetts Institute of Technology, 2009.

[Ziegler 2009] Ziegler B., Methods for Bibliometric Analysis of Research: Renewable Energy Case Study. Composite Information Systems Laboratory (CISL), Massachusetts Institute of Technology, 2009.

[Zhu and Porter 2002] Zhu, D., and Porter, A.L. Automated extraction and visualization of information for technological intelligence and forecasting, Technological Forecasting and Social Change 69, pp. 495–506, 2002.

APPENDIX 1: INFORMATION ABOUT ONLINE PUBLICATION DATA SOURCES

This appendix describes the databases that are used to collect terms and hit counts for the purposes of early growth. The sources that will be analyzed are ACM, Compendex, Google Scholar, IEEE Explore, IngentaConnect, Inspec, Scirus, SpringerLink, and Web of Science.

ACM

The ACM Digital Library is an extensive collection of all of ACM's journals, magazines, peer-reviewed articles, conference proceedings, ACM SIG Newsletters, and multimedia. It contains the largest full-text archive of articles on computing. This archive contains over two million pages of text, with full-text articles from ACM publications dating back to the1950s, and third-party content with selected archives. 20,000 New full-text articles added each year with 34 Special Interest Groups contributing content. It currently has:

2.0+ Million pages of full-text articles
260,000 Articles
45+ High-impact journals
270+ Conference proceeding titles
2,000+ Conference proceedings volumes
6 ACM magazines (including the flagship Communications of the ACM)
800+ Multimedia files containing audio,video, and more[14].

Compendex

Scope of coverage: Compendex contains a compilation of comprehensive engineering literature databases available for engineers. It currently has 11.3 million records, with over 650,000 new ones added annually, across 190 engineering disciplines gathered from 1970 to the present. 98% of the top 50 U.S. engineering schools currently subscribe to Compendex. New information is gathered weekly from engineering conferences, journals and trade magazines from over 55 countries. Every entry is indexed according to the Engineering Index Thesaurus and indexed according to the precise engineering discipline.

Compendex covers topics from several engineering disciplines, including:
Chemical Engineering (15% of Compendex content)
Civil Engineering (15% of Compendex content)
Mining Engineering (12% of Compendex content)
Mechanical Engineering (12% of Compendex content)
Electrical Engineering (35% of Compendex content)
General Engineering (12% of Compendex content)[15]

Keyword/indexing system:

- Controlled terms: keywords related to the article coming from a list of controlled vocabulary composed of agreed-upon technical terms made by the compilers of the database.
- Uncontrolled terms: uncontrolled vocabulary indexing containing terms not in the controlled vocabulary list

[14] Source: http://portal.acm.org/dl.cfm

[15] Source: http://www.ei.org/compendex

Search Customization Options:

- Document type: journal articles, conference articles/proceedings, monograph chapters/reviews, report chapters/reviews, and dissertations.
- Treatment type: application, biographical, economic, experimental, general review, historical, literature review, management aspects, numerical, or theoretical
- Language
- Publication Year

Google Scholar

Scope of coverage: Google Scholar provides peer-reviewed papers, theses, books, abstracts and articles, from academic publishers, professional societies, preprint repositories, universities and other scholarly organizations. Note however that Google Scholar is specifically focused on scholarly documents, and everything covered by Google Scholar is also covered by Google[16].

Search Customization Options:

- Author
- Publication Field
- Date of Publication
- Subject Area

Keyword/indexing system: Google full-text algorithm

IEEE Xplore

IEEE Xplore is an online resource for accessing scientific and technical publications produced by the Institute of Electrical and Electronics Engineers (IEEE) and its publishing partners. IEEE Xplore provides access to a comprehensive collection of full-text PDF documents comprising the world's most highly cited journals in electrical engineering, computer science, and electronics. The content repository supporting IEEE Xplore contains more than 2 million articles from over 12,000 publications that encompass journals, conference proceedings, and technical standards, with select content dating back to 1893.

IEEE Xplore provides access to content from other publishers through the CrossRef Search feature. IEEE Xplore also facilitates federated searching of major science and technology society digital libraries via its integrated Scitopia® search feature.In addition to journals, conference proceedings, and technical standards content, IEEE Xplore provides access to the IEEE Press book collection[17].

IngentaConnect

IngentaConnect is a website that hosts scholarly books and journals from a range of different publishers. IngentaConnect provides researchers with a comprehensive collection of citation data - some 4 million articles from 11,000 publications and online access to the full text of electronic articles, through online purchase of individual articles, or through subscriptions to publications.[18]

Feature-rich online content offers:

- Reference linking
- Forward citation linking

[16] Source: http://scholar.google.com/

[17] Source: http://ieeexplore.ieee.org/Xplorehelp/Help_Welcome_to_IEEE_Xplore.html

[18] Source: http://www.ingentaconnect.com/

- Supplementary Data
- FastTrack articles (pre-publication)

Inspec

Scope of coverage:

Inspec is an abstracting and indexing database for physics, electrical engineering, electronics and computer science information. Updated weekly, it currently has 11 million specially-selected records gathered from 1969 to the present that are precise, targeted and relevant, with over 600,000 new records added annually[19]. Content is available for:

Physics (47% of Inspec content)
Electronics & Electrical Engineering (26% of Inspec content)
Computing & Control (20% of Inspec content)
Manufacturing & Production Engineering (5% of Inspec content)
Information Technology, Networking and Security (2% of Inspec content)

Keyword/indexing system:

- Controlled terms: keywords related to the article coming from a list of controlled vocabulary composed of agreed-upon technical terms made by the compilers of the database.
- Uncontrolled terms: uncontrolled vocabulary indexing containing terms not in the controlled vocabulary list

Search Customization Options:

- Document type: journal articles, conference articles/proceedings, monograph chapters/reviews, report chapters/reviews, and dissertations.
- Treatment type: application, biographical, economic, experimental, general review, historical, literature review, management aspects, numerical, or theoretical
- Language
- Publication Year

Scirus

Scope of coverage: Scirus contains scientific topics in found web sites, news, journals, web articles and academic papers. It searches over 485 million science-specific web pages, filtering out non-scientific sites, and finds peer-reviewed articles such as pdf and postscript files. Scirus searches the most comprehensive combination of web information, preprint servers, digital archives, repositories and patent and journal databases.

Scirus currently covers over web pages including156 million .edu sites, 54 million .org sites, 9 million .ac.uk sites, 52 million .com sites, 36 million .gov sites, and over 143 million other relevant STM and University sites from around the world[20].

Scirus also indexes these sources (the numbers are approximate):

447,000 articles from American Physical Society
536,000 e-prints from ArXiv.org
42,000 full-text articles from BioMed Central
19,000 documents from Caltech Coda
3,300 e-prints from Cogprints

[19] Source: http://www.ei.org/inspec_inspecarchive
[20] Source: http://www.scirus.com/srsapp/aboutus/

81,800 full-text articles from Crystallography Journals Online
24,000 documents from CURATOR
2.1 million documents from Digital Archives
24,000 documents from DiVa
98,500 full-text articles from Project Euclid
3,200 documents from HKUST Institutional Repository
56,000 documents from The University of Hong Kong
12,700 full-text documents available from IISc
11,000 full-text documents available from Humboldt Universität
284,000 full-text articles from Institute of Physics Publishing
23.1 million patent data from LexisNexis
16,000 full-text articles from Maney Publishing
40,000 full-text documents from MD Consult
585,000 full-text documents from Nature Publishing Group
18.1 million Medline citations via PubMed
72,000 documents from MIT OpenCourseWare
24,700 technical reports from NASA
792,000 full-text theses and dissertations via NDLTD
8,900 documents from Organic Eprints
1,690 documents from PsyDok
1.5 million articles from PubMed Central
738,000 documents from RePEc
63,000 full-text articles from Royal Society Publishing
619,000 full-text articles from SAGE Publications
8.0 million full-text articles from ScienceDirect
463,000 full-text journal articles from Scitation
9,100 articles from SIAM
16,600 documents from University of Toronto T-Space
21,800 full-text documents from WaY.

Keyword/indexing system: 'Keywords' can be seen in the 'refine your search' box in the lower left side of the website. Scirus uses an automated extraction algorithm to calculate ranking by relevance. This ranking is determined by two basic values:

- Words - the location and frequency of a search term within a result account for one half of the algorithm. This is known as static ranking.
- Links - the number of links to a page account for the second half of the algorithm - the more often a page is referred to by other pages, the higher it is ranked. This is known as dynamic ranking. Overall ranking is the weighted sum of the static and dynamic rank values. Scirus does not use metatags, as these are subject to ranking-tweaking by users.

Search Customization Options:

- Information Types: abstracts, articles, books, company homepages, conferences, patents, preprints, scientist homepages, theses/dissertations
- Content Sources
- Subject Areas

Springerlink

Springerlink covers topics in Architecture, Life Science, Behavior Science, Business/Econ, Chemistry/Materials, Computer Science, Environmental Science, Engineering, Humanities, Social Science, Law, Math/Statistics, Medicine, Physics, Astronomy, and Applied Computing from journals, books, reference works, protocols, academic publications. It contains over 1,750 peer reviewed journals and 27,000 eBooks online. 3,500 eBooks, eReference Works and eBook Series titles are scheduled to be added each year.

Keywords are pulled from publishers, some articles have none. Search uses frequency analysis as well as keywords. Search customization options include the title, the author, the editor, ISSN / ISBN / DOI and the date of publication[21].

Web of Science

Web of Science has authoritative, multidisciplinary content that covers over 10,000 of the highest impact journals worldwide, including Open Access journals and over 110,000 conference proceedings. Topics in agriculture, biological sciences, engineering, medical and life sciences, physical and chemical sciences, anthropology, law, library sciences, architecture, dance, music, film, and theater with coverage available to 1900. It contains articles, proceedings, papers, reviews, editorials, news[22].

Web of Science offers access to six comprehensive citation databases:

- Science Citation Index Expanded: Over 7,100 major journals across 150 disciplines, to 1900.
- Social Sciences Citation Index: Over 2,474 journals across 50 social science disciplines, as well as 3,500 of the world's leading scientific and technical journals, to 1956.
- Arts & Humanities Citation Index: Over 1,395 arts and humanities journals, as well as selected items from over 6,000 scientific and social sciences journals.
- Conference Proceedings Citation Index: Over 110,000 journals and book-based proceedings in two editions: Science and Social Science and Humanities, across 256 disciplines.
- Index Chemicus: Over 2.6 million compounds, to 1993.
- Current Chemical Reactions: Over one million reactions, to 1986, plus INPI archives from 1840 to 1985.

Search Customization Options:

- Topic
- Title
- Author
- Publication Name
- Year Published
- Address
- Time past since publication

[21] Source: http://www.springer.com/e-content?SGWID=0-113-12-286799-0

[22] Source: http://wos.isitrial.com/help/helpdefs.html

APPENDIX 2: HIT AGGREGATOR EXCEL VERSION

This Appendix starts by introducing the capabilities of Excel in pulling data from the Web, and continues to explain how to create a hit aggregator application in Excel.

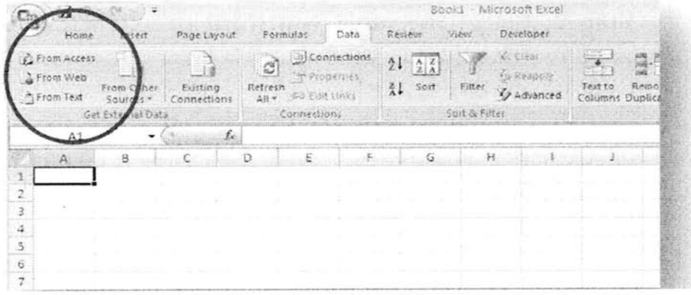
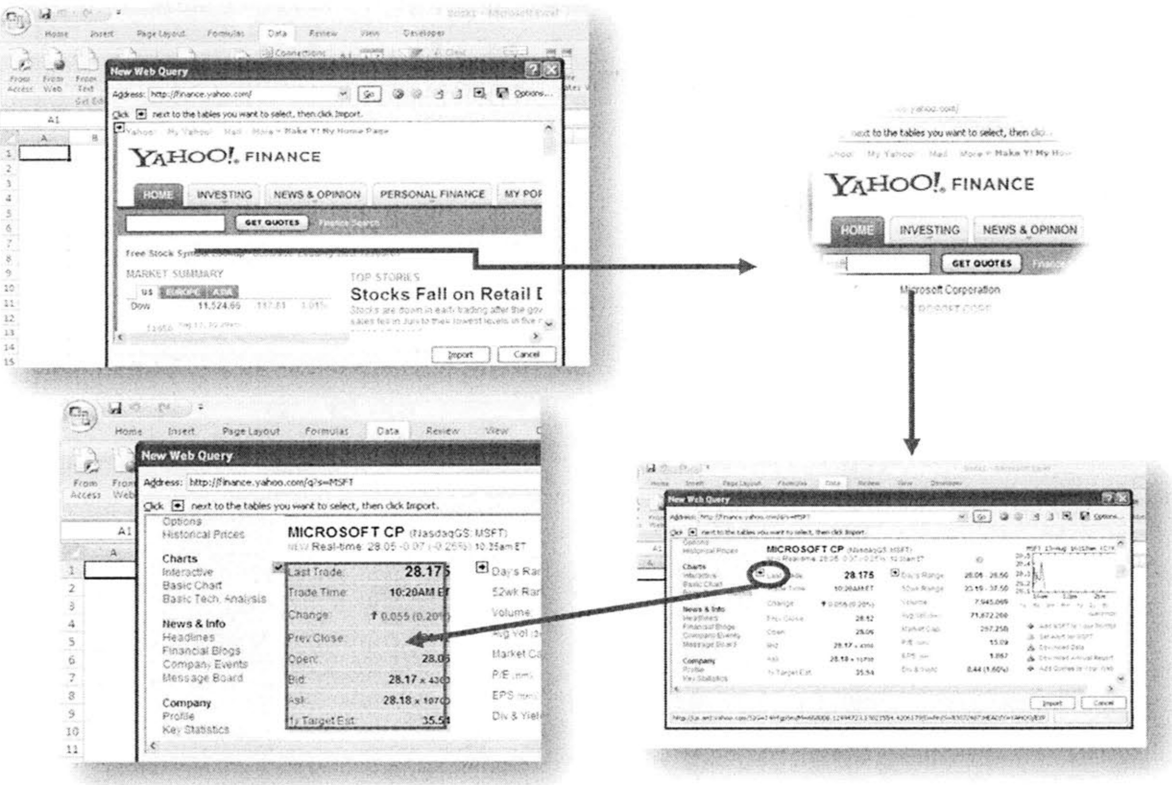


Figure A1.1: Excel allows users to pull data from Access, Web, Text & other sources.

Pulling Data into MS Excel using Web Queries:

MS Excel 2007 allows users to use a Web query to retrieve refreshable data that is stored on the Internet, such as a single table, multiple tables, or all of the text on a Web page. For example, you can retrieve and update stock quotes from a public Web page or retrieve and update a table of sales information from a company Web page.

Web queries are especially useful for retrieving data that is in tables or preformatted areas. (Tables are defined with the HTML <TABLE> tag. Preformatted areas are often defined with the HTML <PRE> tag.) The retrieved data does not include pictures, such as .gif images, and does not include the contents of scripts.To create a Web query, the user needs access to the World Wide Web (WWW). Below is a step by step guide to import data into Excel from Yahoo! Finance (http://finance.yahoo.com).


New Web Query

Address: http://finance.yahoo.com/

Click ▣ next to the tables you want to select, then click Import.

Address: http://finance.yahoo.com/q?s=MSFT

MICROSOFT CP (NasdaqGS: MSFT)

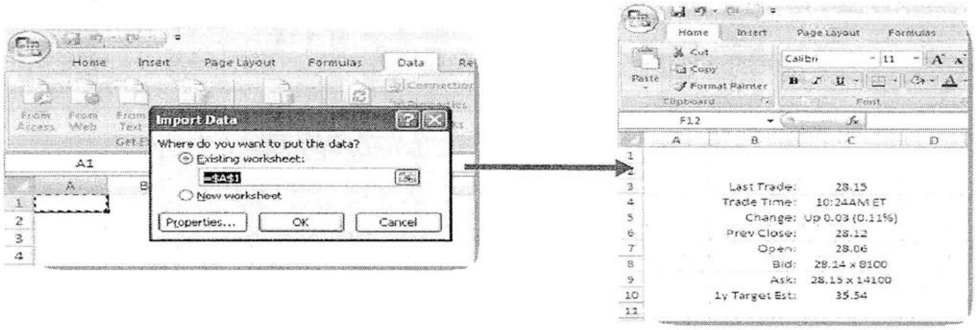
Last Trade:	28.175
Trade Time:	10:20AM ET
Change:	0.055 (0.20%)
Prev Close:	
Open:	
Bid:	28.17 x 4300
Ask:	28.18 x 1070
1y Target Est:	35.54

Figure A1.2: Importing data into Excel from Yahoo Finance.

Step 1: Go to Data > From Web > New Web Query... Enter the Web address for the page you want to use in the Address box (such as Yahoo! Finance - http://finance.yahoo.com), hit Enter or click Go to load the page (in the Yahoo example, enter a company name to get quotes) as in Figure A1.2.

Step 2: Select the table you want to extract: When the page appears in the New Web Query window, Excel adds yellow arrow boxes next to every table you can import. As you hover over each arrow box with the mouse, Excel draws a bold blue outline around the related table.Once you find the table you want to extract, click the arrow box (which then changes into a green checkmark). To deselect a table, just click it again.

Step 3: When you've finished selecting all the tables you want, click the Import button at the bottom of the New Web Query window. Select where you want to put the data. You can do this for any static web site with tables in it. Once you click OK, Excel begins to fetch the information it needs. During this time, you'll see an information message appear in your worksheet (...Getting data...) Excel then replaces this message with the downloaded data, as shown in Figure A1.3.



Where do you want to put the data?

● Existing worksheet:

=$A$1

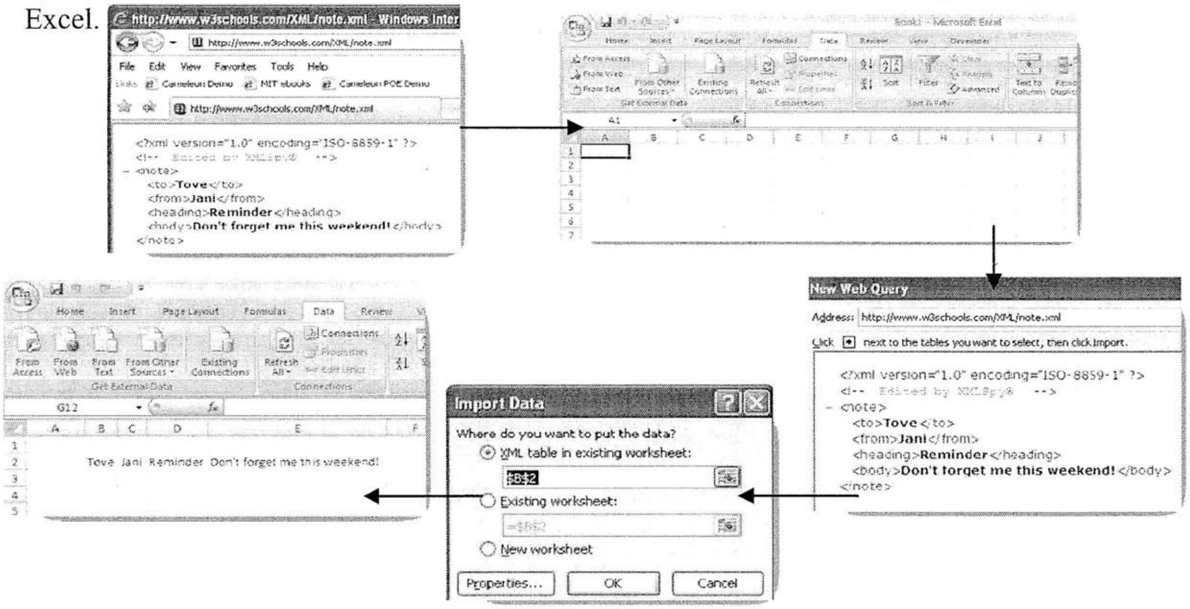
○ New worksheet

Last Trade:	28.15
Trade Time:	10:24AM ET
Change:	Up 0.03 (0.11%)
Prev Close:	28.12
Open:	28.06
Bid:	28.14 x 8100
Ask:	28.15 x 14100
1y Target Est:	35.54

Figure A1.3: Retrieving data that is in tables or preformatted areas.

Working with XML Files

Similarly, a user can also make query against XML Files. Excel will create a schema based on the XML source data. As in the example in Figure A1.4, by clicking Data > From Web > New Web Query, and entering the address of the XML file on the left, the data can easily be imported to MS Excel.



```
<?xml version="1.0" encoding="ISO-8859-1" ?>
<!-- Edited by XMLSpy® -->
- <note>
    <to>Tove</to>
    <from>Jani</from>
    <heading>Reminder</heading>
    <body>Don't forget me this weekend!</body>
  </note>
```

Address: http://www.w3schools.com/XML/note.xml

Click ▣ next to the tables you want to select, then click Import.

Where do you want to put the data?

● XML table in existing worksheet:

$B$2

○ Existing worksheet:

=$B$2

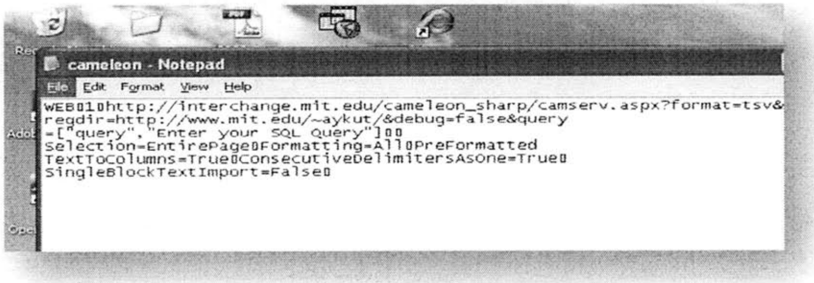
○ New worksheet

Tove Jani Reminder Don't forget me this weekend!

Figure A1.4: A user can also make queries against XML files.

Calling Cameleon from Excel:

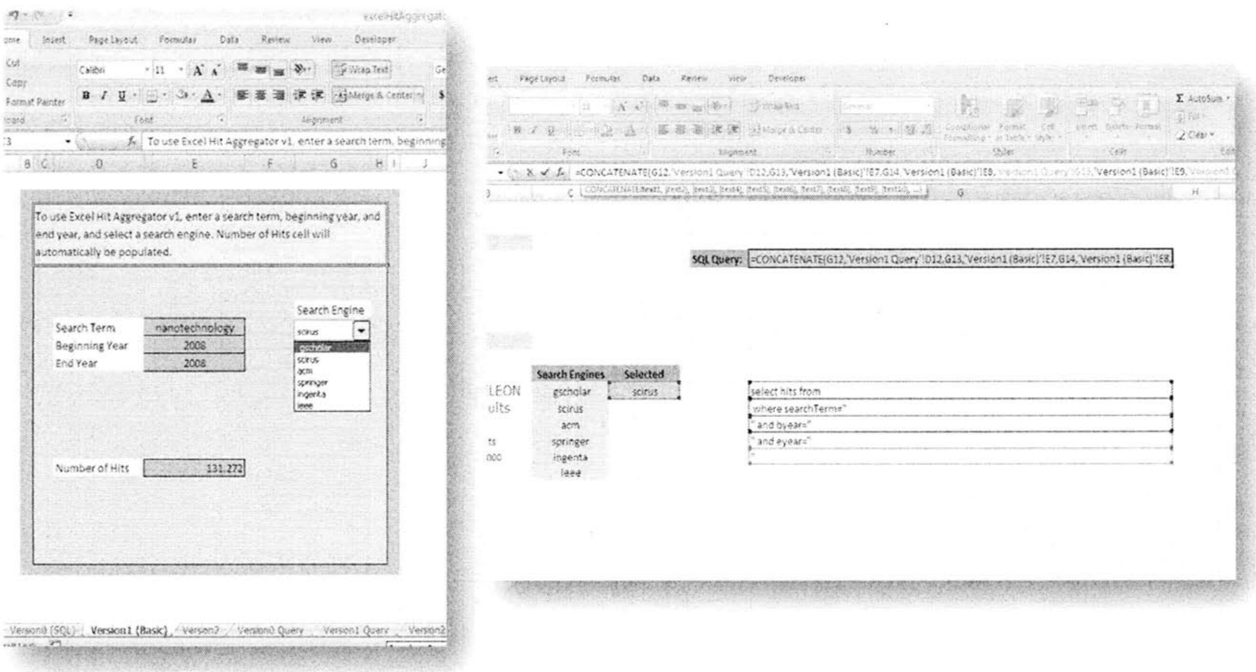
First step calling Cameleon in Excel is to download or create the web query file for Cameleon (cameleon.iqy) (Figure A1.5). A web query file is a text file where each line of text is separated by a carriage return. Web query files can be created in any text editor, such as Notepad, and they are saved with the .iqy extension.



```
WEB□1□http://interchange.mit.edu/cameleon_sharp/camserv.aspx?format=tsv&
regdir=http://www.mit.edu/~aykut/&debug=false&query
=["query","Enter your SQL Query"]□□
Selection=EntirePage□Formatting=All□PreFormatted
TextToColumns=True□ConsecutiveDelimitersAsOne=True□
SingleBlockTextImport=False□
```

Figure A1.5: Cameleon.iqy

The rest will be explained by using the Excel Hit Aggregator, a tool that allows users to get the number of hits (the number of academic papers) for a specific search term from several academic search engines. Below is a screen shot of Excel Hit Aggregator v1 and its query sheet .



To use Excel Hit Aggregator v1, enter a search term, beginning year, and end year, and select a search engine. Number of Hits cell will automatically be populated.

Search Term	nanotechnology
Beginning Year	2008
End Year	2008
Number of Hits	131,272

Search Engine: scirus
gscholar
scirus
acm
springer
ingenta
ieee

SQL Query: =CONCATENATE(G12,'Version1 Query'!D12,G13,'Version1 (Basic)'!E7,G14,'Version1 (Basic)'!E8,

=CONCATENATE(G12,'Version1 Query'!D12,G13,'Version1 (Basic)'!E7,G14,'Version1 (Basic)'!E8,'Version1 Query'!G13,'Version1 (Basic)'!E9,

CONCATENATE(text1, [text2], [text3], [text4], [text5], [text6], [text7], [text8], [text9], [text10], ...)

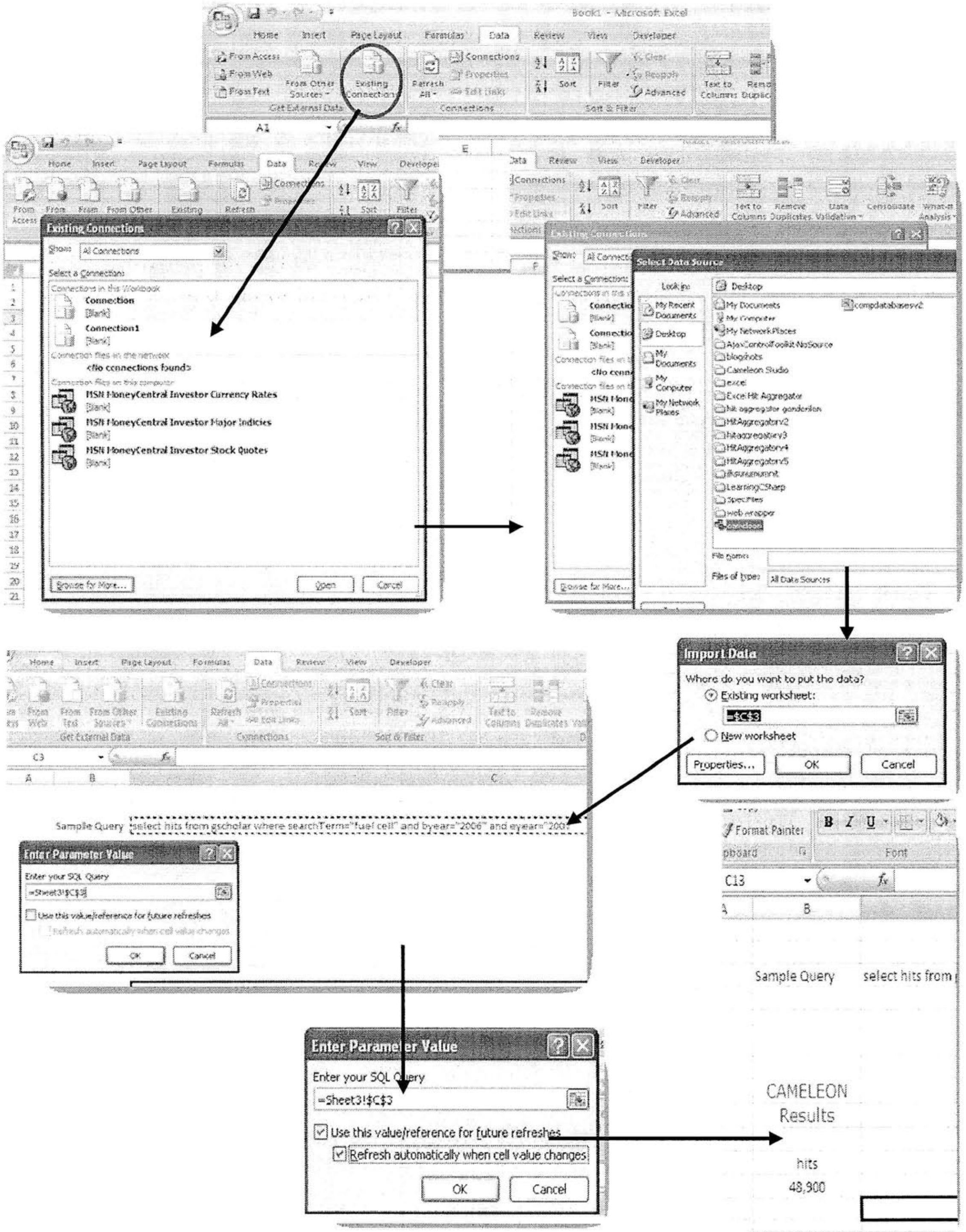
Search Engines	Selected
gscholar	scirus
scirus	
acm	
springer	
ingenta	
ieee	

select hits from
where searchTerm="
" and byear="
" and eyear="
"

Figure A1.6: Excel Hit Aggregator v1 and its query sheet.

To create Excel Hit Aggregator v1:

Step 1: Creating a dynamic query: As in Figure A1.6, write the static portions of your SQL query in Excel cells (such as "select hits from", "where searchTerm=", etc). Then use the concatenate function in



Existing Connections

Show: All Connections

Select a Connection:

Connections in this Workbook
- Connection (Blank)
- Connection1 (Blank)

Connection files on the network
- <No connections found>

Connection files on this computer
- MSN MoneyCentral Investor Currency Rates (Blank)
- MSN MoneyCentral Investor Major Indicies (Blank)
- MSN MoneyCentral Investor Stock Quotes (Blank)

Select Data Source

Look in: Desktop

My Documents, My Computer, My Network Places, AjaxControlToolkit-NoSource, blogshots, Cameleon Studio, excel, Excel Hit Aggregator, hit aggregator gonderilen, HitAggregatorv2, hitaggregatorv3, HitAggregatorv4, HitAggregatorv5, ikusunumunit, LearningCSharp, SpecFiles, Web wrapper, compdatabasesv2

Files of type: All Data Sources

Import Data

Where do you want to put the data?

● Existing worksheet:

=$C$3

○ New worksheet

Sample Query select hits from gscholar where searchTerm="fuel cell" and byear="2006" and eyear="200

Enter Parameter Value

Enter your SQL Query

=Sheet3!$C$3

□ Use this value/reference for future refreshes

□ Refresh automatically when cell value changes

Enter Parameter Value

Enter your SQL Query

=Sheet3!$C$3

☑ Use this value/reference for future refreshes

☑ Refresh automatically when cell value changes

Sample Query select hits from

CAMELEON Results

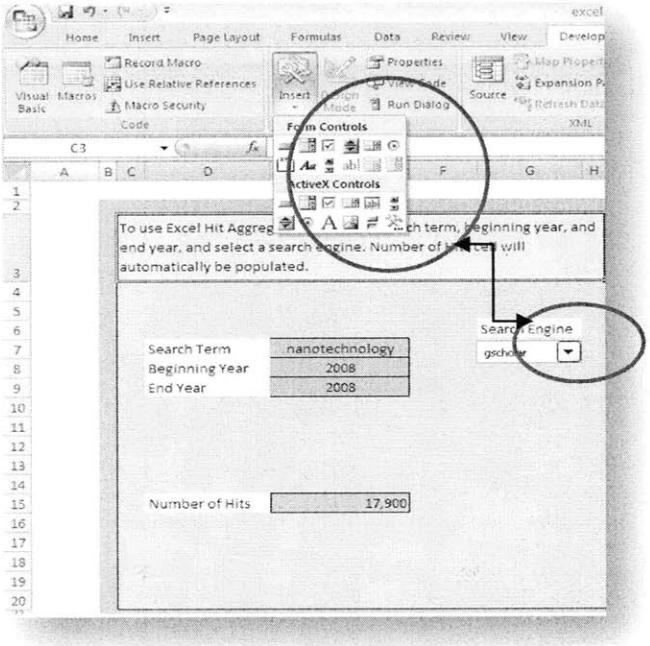
hits
48,900

Figure A1.7: Calling Cameleon from Excel

Excel, to combine the static portions of the query with the parameters the user will enter (Figure A1.6, left).

Step 2: Calling Cameleon: Use Data > Existing Connections > Browse for More. Choose cameleon.iqy. Choose the cell you put your dynamic query in Step 1. Click "Use this value/reference for future refreshes" and "Refresh automatically when cell value changes". You will have Cameleon Results in Excel (Figure A1.7.)

We used Excel developer form controls to add more functionality to our tool. (Figure A1.8)



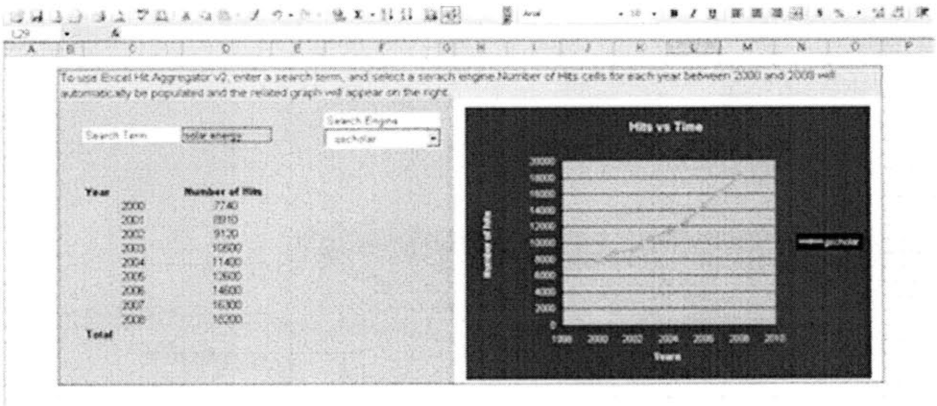
To use Excel Hit Aggreg ch term, beginning year, and end year, and select a search engine. Number of Hits cell will automatically be populated.

Search Term	nanotechnology
Beginning Year	2008
End Year	2008
Number of Hits	17,900

Search Engine: gscholar

Figure A1.8: Using Form Controls in Excel Developer

We have also created Excel Hit Aggregator v2 which uses of the Excel graph capability (Figure A1. 9) and v0 where the end user enters the query (Figure A1.11). You can also see v2 query sheet in Figure A1.10.



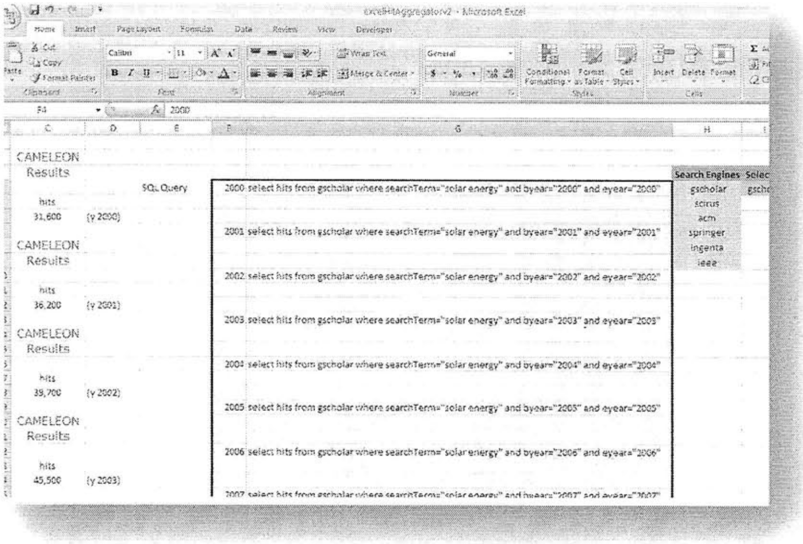
To use Excel Hit Aggregator v2, enter a search term, and select a serach engine Number of Hits cells for each year between 2000 and 2008 will automatically be populated and the related graph will appear on the right

Search Term: solar energy

Search Engine: gscholar

Year	Number of Hits
2000	7740
2001	8910
2002	9120
2003	10600
2004	11400
2005	13600
2006	14600
2007	16300
2008	18200
Total	

Figure A1.9: Excel Hit Aggregator v2



CAMELEON Results

hits 31,600 (y 2000)

CAMELEON Results

hits 36,200 (y 2001)

CAMELEON Results

hits 39,700 (y 2002)

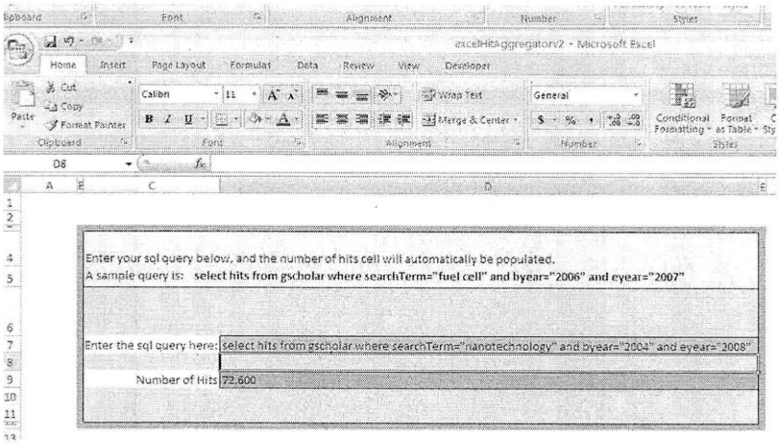
CAMELEON Results

hits 45,500 (y 2003)

SQL Query

```
2000 select hits from gscholar where searchTerm="solar energy" and byear="2000" and eyear="2000"
2001 select hits from gscholar where searchTerm="solar energy" and byear="2001" and eyear="2001"
2002 select hits from gscholar where searchTerm="solar energy" and byear="2002" and eyear="2002"
2003 select hits from gscholar where searchTerm="solar energy" and byear="2003" and eyear="2003"
2004 select hits from gscholar where searchTerm="solar energy" and byear="2004" and eyear="2004"
2005 select hits from gscholar where searchTerm="solar energy" and byear="2005" and eyear="2005"
2006 select hits from gscholar where searchTerm="solar energy" and byear="2006" and eyear="2006"
2007 select hits from gscholar where searchTerm="solar energy" and byear="2007" and eyear="2007"
```

Figure A1.10: Excel Hit Aggregator v2 query sheet



Enter your sql query below, and the number of hits cell will automatically be populated.

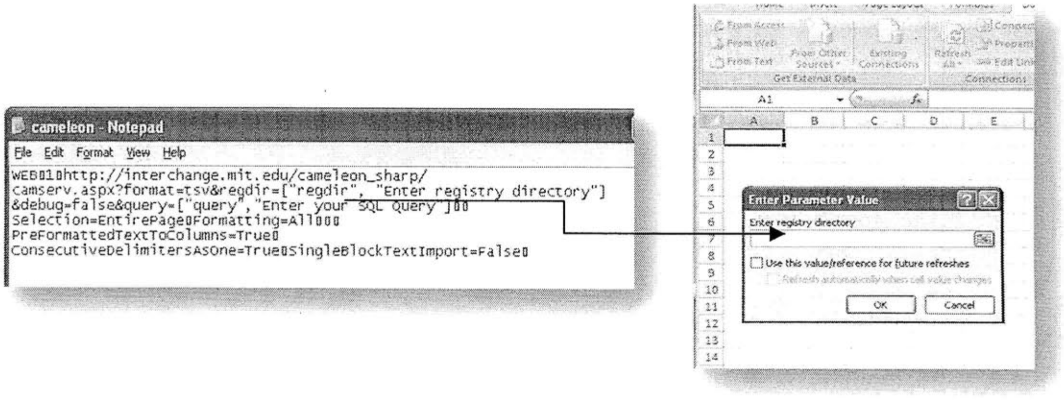
A sample query is: select hits from gscholar where searchTerm="fuel cell" and byear="2006" and eyear="2007"

Enter the sql query here: select hits from gscholar where searchTerm="nanotechnology" and byear="2004" and eyear="2008"

Number of Hits 72,600

Figure A1.11: Excel Hit Aggregator v0

Changing Registry Directory: You can create dynamic queries by replacing the value of the parameter in the web query file with:["paramname","Enter the value for paramname:"]. In cameleon.iqy, instead of using (**regdir**=http://www.mit.edu/~aykut/), you can write (regdir=["regdir", "Enter registry directory"]. When this modified .iqy file is used, Excel will ask the user to enter registry directory (Figure A1.12)



```
WEB□1□http://interchange.mit.edu/cameleon_sharp/
camserv.aspx?format=tsv&regdir=["regdir", "Enter registry directory"]
&debug=false&query=["query","Enter your SQL Query"]□□
Selection=EntirePage□Formatting=All□□□
PreFormattedTextToColumns=True□
ConsecutiveDelimitersAsOne=True□SingleBlockTextImport=False□
```

Enter registry directory

☐ Use this value/reference for future refreshes

☐ Refresh automatically when cell value changes

Figure A1.12: Changing Registry Directory

APPENDIX 3- CAMELEON SCHEDULER

User management

The system currently has a very simple user management scheme that is specified inside the Web.config file as follows:

```
<credentials passwordFormat="Clear">
          <user name="coin@mit.edu" password="rombutan"/>
</credentials>
```

New user names can be added by simply adding new <user> tag inside the credentials.

Setting up the email server

The smtp server needed for sending emails is specified in the Web config file under application settings and currently set to outgoing.mit.edu.

```
<appSettings>
    <add key="UploadFolder"  value="~/UploadedFiles/" />
    <add key="smtpServer"  value="outgoing.mit.edu" />
    <add key="AllowedFileExtensions" value=".csv,.txt"/>
  </appSettings>
File List
<cameleonscheduler>
    Default.aspx
    Default.aspx.cs
    Login.aspx
    Tasks.aspx
    Tasks.aspx.cs
    Web.config
   <App_Code>
         Task.cs
   <App_Data>
         Taskaccessdb.mdb
   <ResultFiles>
   <UploadedFiles>
```

APPENDIX 4 – EGTA TOOL INSTALLATION INSTRUCTIONS

Instructions for setting up the Desktop Application

1. Download the zip file LongTailDeskTopApp.zip
2. Unzip the files
3. Go to \LongTailApp\bin\Debug directory
4. Click on LongTailApp.exe (note that if file extensions are hidden in your folder you should see that this file has Application extension, and has 40KB size. Do not mix it with the similarly named XML configuration and Application Manifest files)
5. The default registry directory is \LongTailApp\bin\Debug\Cameleon#
6. The default Results directory where the result files are stored is LongTailApp\bin\Debug\Results

Instructions for setting up the Web Application

1. Download the zip file from LongTailWebProject.zip
2. Unzip the files
3. Upload all of the files and directory structure to your web server's appropriate directory

APPENDIX 5 - CODE FOR THE EARLY GROWTH TECHNOLOGY ANALYSIS TOOL

1. Model Files

- CameleonAccessor.cs
- CollectionModel.cs
- Collector.cs
- CollectorResultSet.cs
- ErrorListener.cs
- HitCollectionModel.cs
- HitCountCollector.cs
- HitCountCollectorResultSet.cs
- MultiHitCountCollector.cs
- MultiTermCollector.cs
- ProgressListener.cs
- Query.cs
- RemoteCameleonAccessor.cs
- SortHitCountBox.cs
- SortTopTermMeasures.cs
- TermCollectionModel.cs
- TermCollector.cs
- TermCollectorResultSet.cs
- TopTermModel.cs

- **CameleonAccessor.cs**

```
using System;
using System.Collections.Generic;
using System.Text;
using System.Collections;
using System.Collections.Specialized;
using System.Data;
using System.Threading;

    public class CameleonAccessor
    {
        Query query;
        List<ArrayList> results;
        bool useRemote = false;
        string defaultRemote = "http://interchange.mit.edu/cameleon_sharp/camserv.aspx";
        private void checkIfSpecialSourcesExist(Query q) {
            if (q.Source.ToLower().Contains("inspec")
||q.Source.ToLower().Contains("compendex"))
                useRemote = true;
        }

        public CameleonAccessor(Query q) {
            query = q;
            checkIfSpecialSourcesExist(q);
            results = new List<ArrayList>();
        }
        public void setDefaultRemote(string val){
            defaultRemote = val;
        }

        public void setUseRemote(bool val){
            useRemote = val;
        }
        public List<ArrayList>  getResults(){
        if (useRemote)
        {
         RemoteCameleonAccessor remote = new RemoteCameleonAccessor(defaultRemote,
query.ToString(), query.Registry);
                return remote.getResults();
            }
            else
                convertTableResultstoList(getCameleonResults());
            return results;
        }
        private DataTable getCameleonResults(){
                Cameleon.QueryHandler q = new Cameleon.QueryHandler(query.Source,
query.Registry, false, query.Requested, query.Bound);
                try
                {
                    q.setDataTable2();
                    return q.getDataTable();
                }
                catch (Exception ex)
                {

                    //Form1.txtMessageBox.Text = ex.Message;
                }
                return new DataTable();

        }

        protected void convertTableResultstoList(DataTable dt) {

            for (int i = 0; i < dt.Rows.Count; i++){
                ArrayList row = new ArrayList();
                for (int j=0; j < dt.Columns.Count;j++){
                    row.Add((string)dt.Rows[i].ItemArray.GetValue(j));
                }
                results.Add(row);
            }
        }
    }
```

CollectionModel.cs

```
using System;
using System.Collections.Generic;
using System.Linq;
using System.Text;
using System.Collections;
    abstract public class CollectionModel     {
        protected List<string> generatedTerms = new List<string>();
        protected ArrayList sourceList = new ArrayList();
        protected string registry = "";
        protected string message = "";
        protected int maxTermListSize = 80;
        protected int maxThreads = 10;
        protected string seedTerm;
        protected int progress = 0;
        List<ProgressListener> progressListeners = new List<ProgressListener>();
        List<ErrorListener> errorListeners = new List<ErrorListener>();
        public int getUnitProgress(){
            return progress;
        }
        public void registerProgressListener(ProgressListener listener) {
            progressListeners.Add(listener);
        }
        public void removeProgressListener(ProgressListener listener) {
             progressListeners.Remove(listener);
        }

        public void updateProgress(int value) {
            progress = value;
            for (int i = 0; i < progressListeners.Count; i++)
                progressListeners[i].update(progress);
        }
        public void registerErrorListener(ErrorListener listener) {
            errorListeners.Add(listener);        }
        public void removeErrorListener(ErrorListener listener)         {
            errorListeners.Remove(listener);
        }
        public void updateErrors(string message) {
            for (int i = 0; i < errorListeners.Count; i++)
                errorListeners[i].updateError(message);
        }

        public CollectionModel() { }
        public CollectionModel(string seedTerm, ArrayList sourceList, string registry,
int maxThreads, int maxTermListSize) {
            this.seedTerm = seedTerm;
            this.sourceList = sourceList;
            this.registry = registry;
            this.maxTermListSize = maxTermListSize;
            this.maxThreads = maxThreads;
        }
        public List<string> getGeneratedTerms(){
            return generatedTerms;
        }
        public void setGeneratedTerms(List<string> generatedTerms) {
            this.generatedTerms = generatedTerms;
        }
    }
```

Collector.cs

```
using System.Text;
using System.Collections;
using System.Collections.Specialized;
using System.Data;
using System.Threading;


    abstract class Collector
    {
        protected CameleonAccessor cameleon;


        protected string registry;
        protected ArrayList sourceList;



        protected abstract void collectfromaSingleSource(object source);
        protected abstract Query initializeQuery(string source);


        protected List<ArrayList> getCameleonResults(string source)
        {
            cameleon = new CameleonAccessor(initializeQuery(source));

                return cameleon.getResults();


        }
        public void collect()
        {
            Thread[] th = new Thread[sourceList.Count];
            for (int i = 0; i < sourceList.Count; i++)
            {
                th[i] = new Thread(new
ParameterizedThreadStart(collectfromaSingleSource));
                    th[i].Start(sourceList[i]);

            }
            for (int i = 0; i < th.Length; i++)
```

```
                th[i].Join();
        }
        }
```

CollectorResultSet.cs

```
using System;
using System.Collections.Generic;

using System.Text;


    abstract class CollectorResultSet    {
        protected string source;
        public string getSource() { return source; }
    }
```

ErrorListener.cs

```
using System;
using System.Collections.Generic;
using System.Linq;
using System.Web;
public interface ErrorListener{
    void updateError(string message);
}
```

HitCollectionModel.cs

```
using System;
using System.Collections.Generic;
using System.Linq;
using System.Text;
using System.Collections;
using System.IO;



    public class HitCollectionModel : CollectionModel
    {
        List<string[]> generatedHitCounts = new List<string[]>();
        List<string[]> hitCountInput = new List<string[]>();

        public HitCollectionModel(string seedTerm, ArrayList sourceList, string registry,
int maxThreads, int maxTermListSize, List<string[]> hitCountInput) :
            base(seedTerm, sourceList, registry, maxThreads, maxTermListSize)
        {
            this.hitCountInput = hitCountInput;
        }



        private List<string[]> getNextTermYearBatch(int index, int size)
        {
            List<string[]> termYearBatch = new List<string[]>();
            for (int i = index; i < index + size; i++)
                termYearBatch.Add(hitCountInput[i]);
            return termYearBatch;
        }
        private void appendToGeneratedHitCounts(List<string[]> results)
        {

            try
            {
                for (int i = 0; i < results.Count; i++)
                {
                    string[] row = results[i];
                    if (row != null)
                    {
                        row[0] = row[0].Replace(",", "");
                        row[0] = row[0].Replace("\"", "");
                        row[2] = row[2].Replace(",", "");
                        row[2] = row[2].Replace("\"", "");
                        generatedHitCounts.Add(row);
                    }
                }

            }
            catch (Exception e)
            {
                message=e.Message;
                updateErrors(message);
            }
        }
        public void collectResults()
        {
            generatedHitCounts.Clear();
            List<string[]> termYearBatch = new List<string[]>(); ;
            int i = 0;
            for (i = 0; i < hitCountInput.Count - maxThreads; i = i + maxThreads)
            {
                termYearBatch = getNextTermYearBatch(i, maxThreads);
                MultiHitCountCollector multiHitCountCollector = new
MultiHitCountCollector(termYearBatch, sourceList, registry);
                multiHitCountCollector.collect();

appendToGeneratedHitCounts(multiHitCountCollector.getHitCountsAsAnArray());
                updateProgress((int)(100*Math.Min(1,( (double)(i + maxThreads) /
(double)(hitCountInput.Count)))));
            }
            if (i < hitCountInput.Count)
                termYearBatch = getNextTermYearBatch(i, hitCountInput.Count - i);
            MultiHitCountCollector multiHitCountCollector2 = new
MultiHitCountCollector(termYearBatch, sourceList, registry);
            multiHitCountCollector2.collect();
            appendToGeneratedHitCounts(multiHitCountCollector2.getHitCountsAsAnArray());
            updateProgress(100);
            sort();
        }
        private void sort(){
            ArrayList hitCounts = new ArrayList();
            for (int i = 0; i < generatedHitCounts.Count; i++)
            {
                string[] r = generatedHitCounts[i];
                hitCounts.Add(r);
            }
            hitCounts.Sort(new SortHitCountBox());
            generatedHitCounts = new List<string[]>();
            for (int i = 0; i < hitCounts.Count; i++)
            {
                string[] r = (string[])hitCounts[i];
                generatedHitCounts.Add(r);
            }
        }
        public List<string[]> getGeneratedHitCounts()
        {
            return generatedHitCounts;
        }

        public void setHitCountInput(List<string[]> hitCountInput)
        {
            this.hitCountInput = hitCountInput;
        }


    }
```

HitCountCollector.cs

```
using System;
using System.Collections.Generic;

using System.Text;
using System.Collections;
using System.Collections.Specialized;
using System.Data;


    class HitCountCollector : Collector
    {
        protected string term;
        private List<HitCountCollectorResultSet> results;
        private int year;

        public HitCountCollector(string term, ArrayList sourceList, string registry, int
year)
        {
            this.term = term;
            this.sourceList = sourceList;
            this.registry = registry;
            this.year = year;
            results = new List<HitCountCollectorResultSet>();


        }

        protected  override Query initializeQuery(string source)
        {

            ArrayList requested = new ArrayList();
            NameValueCollection bound = new NameValueCollection();
            requested.Add("searchTerm");
            requested.Add("bYear");
            requested.Add("hits");
            bound.Add("searchTerm", term);
            bound.Add("byear", year.ToString());
            bound.Add("eyear", year.ToString());
            return new Query(registry, source, requested, bound);
        }



        protected HitCountCollectorResultSet
convertArrayListResultstoHitCountCollectorResultSet(List<ArrayList> cameleonResults,
string source)
        {
            HitCountCollectorResultSet list = new HitCountCollectorResultSet(source);
            if (cameleonResults.Count>0)
                list.SetResultSet((string)cameleonResults[0][0],
(string)cameleonResults[0][2], (string)cameleonResults[0][1]);
            return list;
        }


        protected  override void collectfromaSingleSource(object source)
        {

                List<ArrayList> cameleonResults = getCameleonResults((string)source);


results.Add(convertArrayListResultstoHitCountCollectorResultSet(cameleonResults,
(string)source));

        }
        public List<HitCountCollectorResultSet> getResults() { return results; }
    }
```

HitCountCollectorResultSet.cs

```
using System;
using System.Collections.Generic;
using System.Text;


    class HitCountCollectorResultSet : CollectorResultSet
    {
        string[] termHitCounts;


        public HitCountCollectorResultSet(string source)
        {
            this.source = source;
            termHitCounts = new string[3];
        }
        public string[] getTermHitCounts() { return termHitCounts; }
        public void SetResultSet(string term, string hitCount, string year)
        {
            termHitCounts[0] = term;
            termHitCounts[1] = hitCount;
            termHitCounts[2] = year;
        }
    }
```

MultiHitCountCollector.cs

```
using System;
using System.Collections.Generic;

using System.Text;
using System.Collections;
using System.Collections.Specialized;
using System.Data;
using System.Threading;


    class MultiHitCountCollector
    {
        List<string[]> termYearList;
        ArrayList sourceList;
        string registry;
        List<HitCountCollectorResultSet> hitCounts;

        public MultiHitCountCollector(List<string[]> termYearList, ArrayList sourceList,
string registry)
        {
            this.termYearList = termYearList;
            this.sourceList = sourceList;
            this.registry = registry;
        }

        private void collectHitCountsFromMultipleSources(object termYear)
        {

            HitCountCollector hitCountCollector = new
HitCountCollector(((string[])termYear)[0], sourceList,
registry,int.Parse(((string[])termYear)[1]));
            hitCountCollector.collect();
            lock (this)
            {
                hitCounts.AddRange(hitCountCollector.getResults());
            }
        }

        public void collect()
        {
            hitCounts = new List<HitCountCollectorResultSet>();
            Thread[] th = new Thread[termYearList.Count];
            for (int i = 0; i < termYearList.Count; i++)
            {
                th[i] = new Thread(new
ParameterizedThreadStart(collectHitCountsFromMultipleSources));
                th[i].Start(termYearList[i]);
            }
            for (int i = 0; i < th.Length; i++)
                th[i].Join();

        }

        public List<HitCountCollectorResultSet> getHitCounts()
        {
            return hitCounts;
        }
        public List<string[]> getHitCountsAsAnArray()
        {
            List<string[]> terms = new List<string[]>();

            for (int i = 0; i < hitCounts.Count; i++)
            {
                HitCountCollectorResultSet t = hitCounts[i];
                terms.Add(t.getTermHitCounts());
            }
            return terms;
        }
    }
```

MultiTermCollector.cs

```
using System;
using System.Collections.Generic;
using System.Text;
using System.Collections;
using System.Collections.Specialized;
using System.Data;
using System.Threading;


    class MultiTermCollector
    {
        ArrayList seedTermList;
        ArrayList sourceList;
        string registry;
        List<TermCollectorResultSet> relatedTerms;
        public string reverseTerm;
        public List<string> termsContainingReverseTerm = new List<string>();

        public MultiTermCollector(ArrayList seedTermList, ArrayList sourceList, string
registry)
        {
            this.seedTermList = seedTermList;
            this.sourceList = sourceList;
            this.registry = registry;
        }

        private bool contains(List<string> results, string term)
        {

            for (int i = 0; i < results.Count; i++)
            {
                string r = results[i];
                if (r.ToLower().Contains(term)) return true;
            }
            return false;
        }

        private void collectRelatedTermsFromMultipleSources(object term)
        {
            TermCollector termCollector = new TermCollector((string)term, sourceList,
registry);
            termCollector.collect();
            List<string> results = termCollector.getResultsAsAList();
            if (reverseTerm!=null)
            if (contains(results,reverseTerm))
termsContainingReverseTerm.Add((string)term);
            lock (this)
            {
                relatedTerms.AddRange(termCollector.getResults());

            }
        }






        public void collect()
        {
            relatedTerms = new List<TermCollectorResultSet>();
            Thread[] th = new Thread[seedTermList.Count];
            for (int i=0; i < seedTermList.Count; i++){
                th[i] = new Thread(new
ParameterizedThreadStart(collectRelatedTermsFromMultipleSources));
                th[i].Start(seedTermList[i]);
            }
            for (int i = 0; i < th.Length; i++)
                th[i].Join();

        }

        public List<TermCollectorResultSet> getRelatedTerms()
        {
            return relatedTerms;
        }
        public List<string> getRelatedTermsAsAnArray()
        {
            List<string> terms = new List<string>();

            for (int i = 0; i < relatedTerms.Count; i++)
            {
                TermCollectorResultSet t = relatedTerms[i];
                terms.AddRange(t.getTerms());
            }
            return terms;
        }

    }
```

ProgressListener.cs

```
using System;
using System.Collections.Generic;
using System.Linq;
using System.Text;


    public interface ProgressListener
    {
        void update(int value);

    }
```

Query.cs

```
using System;
using System.Collections.Generic;
using System.Text;
using System.Collections;
using System.Collections.Specialized;
```

```
    public class Query
    {
        public string Registry;
        public string Source;
        public ArrayList Requested;
        public NameValueCollection Bound;

        public Query(string registry, string source, ArrayList requested,
NameValueCollection bound)
        {
            Registry = registry;
            Source = source;
            Requested = requested;
            Bound = bound;
        }
        public override string ToString()
        {

            string requested = "";
            for (int i = 0; i < Requested.Count; i++)
            {
                if (i!=0)
                requested =","+ Requested[i].ToString();
                else
                    requested = Requested[i].ToString();


            }
            string bound = "";
            for (int i = 0; i < Bound.Count; i++)
            {
                if (i != 0)
                    bound = " AND " + Bound.GetKey(i) + "=\"" + Bound.GetValues(i)[0] +
"\"";

                else
                    bound = Bound.GetKey(i) + "=\"" + Bound.GetValues(i)[0]+"\"";

            }
            return "select " + requested + " from " + Source + " where " + bound;
        }
    }
```

RemoteCameleonAccessor.cs

```
using System;
using System.Collections.Generic;
using System.Linq;
using System.Web;
using System.Net;
using System.Text;
using System.Xml.XPath;
using System.IO;
using System.Collections;

/// <summary>
/// Summary description for RemoteCameleonAccessor
/// </summary>
public class RemoteCameleonAccessor
{
    string remoteServer;
        string query;
    string registry;
    public RemoteCameleonAccessor(string remoteServer, string query, string registry)
    {
        this.remoteServer = remoteServer;
        this.query = query;
        if (!registry.ToLower().StartsWith("http"))
            registry = "http://www.mit.edu/~ayshe/";
        this.registry = registry;

    }
    public List<ArrayList> getResults()
    {
        //This is good for single return queries
        List<ArrayList> results = new List<ArrayList>();
        WebClient wc = new WebClient();
        byte[] bPageData;
        string url = remoteServer + "?query=" + query + "&format=xml&regdir=" + registry;
        bPageData = wc.DownloadData(url);
        UTF8Encoding utf8 = new UTF8Encoding();
        string pageData = utf8.GetString(bPageData);
        XPathDocument doc = new XPathDocument(new StringReader(pageData));
        XPathNavigator nav;
        nav = doc.CreateNavigator();
        XPathExpression expr;
        expr = nav.Compile("/DOCUMENT/ELEMENT/*");
        XPathNodeIterator iterator = nav.Select(expr);

        while (iterator.MoveNext())
        {
            ArrayList row = new ArrayList();
            row.Add(iterator.Current.Value);
            results.Add(row);
        }
        return results;
    }

}
```

SortHitCountBox.cs

```
using System;
using System.Collections.Generic;

using System.Text;




    /// <summary>
    /// Summary description for CMySort
    /// </summary>
    public class SortHitCountBox : System.Collections.IComparer
    {

        public SortHitCountBox()
        {

        }
        public int Compare(object x, object y)
        {
            string[] d1 = (string[])x;
            string[] d2 = (string[])y;
            int comp = string.Compare(d1[0], d2[0], true);
            if (comp == 0)
            {
                if (double.Parse(d1[2]) < double.Parse(d2[2])) return -1;
                else if (double.Parse(d1[2]) == double.Parse(d2[2])) return 0;
                else return 1;
            }
            else return comp;

        }
    }
```

SortTopTermMeasures.cs

```
using System;
using System.Collections.Generic;


/// <summary>
/// Summary description for CMySort
/// </summary>
```

```
public class SortTopTermMeasures : System.Collections.IComparer
{

    public SortTopTermMeasures()
    {

    }
    public int Compare(object x, object y)
    {
        string[] d1 = (string[])x;
        string[] d2 = (string[])y;
        if (double.Parse(d1[1]) < double.Parse(d2[1])) return 1;
        else if (double.Parse(d1[1]) == double.Parse(d2[1])) return 0;
        else return -1;
    }
}
```

TermCollectionModel.cs

```
using System;
using System.Collections.Generic;
using System.Linq;
using System.Text;
using System.Collections;
using System.IO;



    public class TermCollectionModel  : CollectionModel
    {
        List<int> levels = new List<int>();
        const int maxNumberOfEmptyQueryRepeatAttempts = 4;
        int numberOfEmptyQueryRepeatAttempts = 0;
        List<string> termsToExpand = new List<string>();
        List<string> refinedList = new List<string>();

        public TermCollectionModel(string seedTerm, ArrayList sourceList, string
registry, int maxThreads, int maxTermListSize):
            base( seedTerm, sourceList, registry,  maxThreads, maxTermListSize)
        {

            levels.Add(20);
            levels.Add(10);

        }

        public void initialize()
        {
            termsToExpand.Clear();
            generatedTerms.Clear();
            generatedTerms.Add(seedTerm);
            numberOfEmptyQueryRepeatAttempts = 0;
        }
        private ArrayList getNextTermList(int startIndex, int size)
        {
            ArrayList termList = new ArrayList();
            for (int i = 0; i < size; i++)
                termList.Add(generatedTerms[startIndex + i]);
            return termList;
        }
        public void refine()
        {
            generatedTerms.RemoveAt(0);
            for (int startIndex = 0; startIndex < generatedTerms.Count; startIndex +=
maxThreads)
            {
                int n = Math.Min(maxThreads, generatedTerms.Count - startIndex);
                MultiTermCollector multiTermCollector = new
MultiTermCollector(getNextTermList(startIndex,n), sourceList, registry);
                multiTermCollector.reverseTerm = this.seedTerm;
                multiTermCollector.collect();
                updateProgress( (int)(100*((double)(startIndex + maxThreads) /
(double)(generatedTerms.Count))));
                for (int i = 0; i < multiTermCollector.termsContainingReverseTerm.Count;
i++)
                {
                    refinedList.Add(multiTermCollector.termsContainingReverseTerm[i]);
                }
            }
            generatedTerms = new List<string>();
            generatedTerms.Add(seedTerm);
            generatedTerms.AddRange(refinedList);

        }

        public void collectTerms()
        {

            initialize();
            addInitialResultsToGeneratedTerms();
            updateProgress((int)(100*Math.Min(1, ((double)generatedTerms.Count /(double)
maxTermListSize))));
            addRemainingResultsToGeneratedTerms();




        }
        private void addInitialResultsToGeneratedTerms()
        {
            TermCollector termCollector = new TermCollector(seedTerm, sourceList,
registry);
            termCollector.collect();
            generatedTerms.AddRange(termCollector.getResultsAsAList());
            termsToExpand.AddRange(getPermittedByLevel(termCollector.getResultsAsAList(),
0));

        }
        private bool isNotDone(int index)
        {
            if (index < (termsToExpand.Count - maxThreads) && generatedTerms.Count <
maxTermListSize) return true;
            else return false;
        }
        private ArrayList getNextTermListToExpand(int startIndex, int size)
        {

            ArrayList termList = new ArrayList();
            for (int i = 0; i < size && i < termsToExpand.Count - startIndex; i++)
                termList.Add(termsToExpand[startIndex + i]);
            return termList;
        }
        private List<string> getNextResultsFromStartIndex(int startIndex)
        {

            MultiTermCollector multiTermCollector = new
MultiTermCollector(getNextTermListToExpand(startIndex, maxThreads), sourceList,
registry);
            multiTermCollector.collect();
            return multiTermCollector.getRelatedTermsAsAnArray();

        }
        private void appendResults(List<string> nextResults)
        {
            int nadd = 0;
            for (int i = 0; i < nextResults.Count; i++)
                if (!generatedTerms.Contains(nextResults[i]))
                {

                    generatedTerms.Add(nextResults[i]);
                    if (termsToExpand.Count < levels[0] * levels[1])
                        if (nadd < levels[1] * maxThreads)
```

```
                            termsToExpand.Add(nextResults[i]);
                    nadd++;
                }
        }


        private void addRemainingResultsToGeneratedTerms()
        {
            int index = 0;
            while (isNotDone(index))
            {
                List<string> nextResults = getNextResultsFromStartIndex(index);
                if (nextResults.Count < 1 && numberOfEmptyQueryRepeatAttempts <=
maxNumberOfEmptyQueryRepeatAttempts)
                    numberOfEmptyQueryRepeatAttempts++;
                else
                {
                    index += maxThreads;
                    appendResults(nextResults);
                    numberOfEmptyQueryRepeatAttempts = 0;
                }
                updateProgress((int)(100*Math.Min(1,((double)generatedTerms.Count /
(double)maxTermListSize))));

            }

        }
        private List<string> getPermittedByLevel(List<string> list, int level)
        {

            List<string> newList = new List<string>();
            try
            {
                for (int i = 0; i < levels[level]; i++)
                    newList.Add(list[i]);
            }
            catch (Exception ex)
            {
                updateErrors("No rows were returned from Cameleon");
            }
            return newList;
        }

    }
```

TermCollector.cs

```
using System;
using System.Collections.Generic;

using System.Text;
using System.Collections;
using System.Collections.Specialized;
using System.Data;


    class TermCollector : Collector
    {
        protected string seedTerm;
        private List<TermCollectorResultSet> results;
        protected List<string> flattenedResults;



        public TermCollector(string seedTerm, ArrayList sourceList,string registry)
        {
            this.seedTerm = seedTerm;
            this.sourceList = sourceList;
            this.registry = registry;
            results = new List<TermCollectorResultSet>();
            flattenedResults = new List<string>();
        }

        protected override Query initializeQuery(string source)
        {
            ArrayList requested = new ArrayList();
            NameValueCollection bound = new NameValueCollection();
            requested.Add("relatedTerms");
            bound.Add("searchTerm", seedTerm);
            return new Query(registry, source, requested, bound);
        }


        protected TermCollectorResultSet
convertArrayListResultstoTermCollectorResultSet(List<ArrayList> cameleonResults, string
source)
        {
            TermCollectorResultSet list = new TermCollectorResultSet(source, seedTerm);
            for (int i = 0; i <cameleonResults.Count; i++)
                list.Add((string)(cameleonResults[i][0]));
            return list;
        }
        private void setflattenedResults(List<ArrayList> cameleonResults)
        {
            for (int i = 0; i < cameleonResults.Count; i++)
                flattenedResults.Add((string)cameleonResults[i][0]);
        }


        protected override void collectfromaSingleSource(object source)
        {
            List<ArrayList> cameleonResults;
            cameleonResults = getCameleonResults((string)source);
            setflattenedResults(cameleonResults);
            results.Add(convertArrayListResultstoTermCollectorResultSet(cameleonResults,
(string)source));
        }

        public List<string> getResultsAsAList()
        {
            return flattenedResults;
        }

        public List<TermCollectorResultSet> getResults() { return results; }

    }
```

TermCollectorResultSet.cs

```
using System;
using System.Collections.Generic;

using System.Text;
using System.Collections;



    class TermCollectorResultSet : CollectorResultSet
    {
        List<string> terms;
        string seedTerm;

        public TermCollectorResultSet(string source, string seedTerm)
        {
            terms = new List<string>();
            this.source = source;
            this.seedTerm = seedTerm;
        }
        public List<string> getTerms() { return terms; }
        public string getSeedTerm() { return seedTerm; }
        public void Add(string term) { terms.Add(term); }
    }
```

TopTermModel.cs

```
using System;
using System.Collections.Generic;
using System.Linq;
using System.Text;
using System.Collections;


    public class TopTermModel: CollectionModel
    {
        private ArrayList hitCounts;
        private int bYear;
        private int eYear;
        public ArrayList logs;
        public TopTermModel(List<string[]> hitCountList, int bYear, int eYear)
        {
            hitCounts = new ArrayList();
            for (int i=0; i<hitCountList.Count;i++)
            {
                 string[] r = hitCountList[i];
                 hitCounts.Add(r);
            }
            hitCounts.Sort(new SortHitCountBox());

            this.bYear = bYear;
            this.eYear = eYear;

        }
        public ArrayList getRanks()
        {
            return logs;
        }



        public void calculateRanksUsingLogRatio()
        {
            logs = new ArrayList();
            int dif = (eYear-bYear);
            for (int i=0; i <hitCounts.Count-dif;i+=(dif+1))
            {

               string[] r1=(string[])hitCounts[i];
               string[] r2=(string[])hitCounts[i+dif];
                //put error handling
               double hitsB = (double)(int.Parse(r1[1].Replace(",","")));
               if (hitsB == 0) hitsB = 0.00001;
               double hitsE = (double)(int.Parse(r2[1].Replace(",", "")));
               string[] lograrank = { r1[0], (Math.Log(hitsE) / Math.Log(hitsB)).ToString()
};

               logs.Add(lograrank);

            }
            logs.Sort(new SortTopTermMeasures());


        }

       public void calculateRanksUsingPercentage()
        {
            logs = new ArrayList();
            int dif = (eYear - bYear);
            for (int i = 0; i < hitCounts.Count - dif; i+=(dif+1))
            {

                string[] r1 = (string[])hitCounts[i];
                string[] r2 = (string[])hitCounts[i + dif];
                double hitsB = (double)(int.Parse(r1[1].Replace(",", "")));
                if (hitsB == 0) hitsB = 0.00001;
                double hitsE = (double)(int.Parse(r2[1].Replace(",", "")));
                string[] logrank = { r1[0], ((hitsE - hitsB)/hitsB).ToString() };
                logs.Add(logrank);

            }
            logs.Sort(new SortTopTermMeasures());

        }
    }
```

DeskTopApplication Controller
LongTailController.cs

```
using System;
using System.Collections.Generic;
using System.Linq;
using System.Text;


namespace LongTailApp.Controller
{
    class LongTailController : ProgressListener
    {
        protected Form1 view;
        TopTermModel topTermModel;
        HitCollectionModel hitCollectionModel;
        TermCollectionModel termModel;
        public LongTailController(Form1 appForm)
        {
            view = appForm;

        }
        public void update(int value)
        {
            view.updateProgressBar(value);
        }
        public void getTermsClickedEvent()
        {
            view.resetProgressBar();
            view.clearDisplayBox("term");
            view.clearElapsedTimeLabel("term");
            view.disableButtonsNextToTerm();
            termModel =
                new TermCollectionModel(view.getSeedTerm(), view.getTermSourceList(),
view.getRegistry(), view.getMaxThreads(), view.getMaxTermListSize());
            termModel.registerProgressListener(this);
            System.DateTime startTime = System.DateTime.Now;

            view.setTermModel(termModel);
            termModel.collectTerms();
            view.showElapsedTimeinaLabel("term", (System.DateTime.Now -
startTime).TotalSeconds.ToString());
            view.updateTermListBox();
            view.updateMessage("Total of " +
Math.Min(termModel.getGeneratedTerms().Count, view.getMaxTermListSize()) + " terms");
            if (termModel.getGeneratedTerms().Count > 0)
            view.enableButtonsNextToTerm();
            if (view.isSaveChecked())
                view.writeTermListResults();
            view.updateProgressBar(100);
        }
        public void getHitCountsClickedEvent()
        {
            view.resetProgressBar();
            view.clearDisplayBox("hit");
            view.clearElapsedTimeLabel("hit");
            view.disableButtonsNextToHit();
            hitCollectionModel =
                new HitCollectionModel(view.getSeedTerm(), view.getHitSourceList(),
view.getRegistry(), view.getMaxThreads(), view.getMaxTermListSize(),
view.getHitCountInput());
            hitCollectionModel.registerProgressListener(this);
            System.DateTime startTime = System.DateTime.Now;
            view.setHitModel(hitCollectionModel);
            hitCollectionModel.collectResults();
            view.showElapsedTimeinaLabel("hit", (System.DateTime.Now -
startTime).TotalSeconds.ToString());
            view.updateHitCountListBox();
            if (hitCollectionModel.getGeneratedHitCounts().Count > 1)
                view.enableButtonsNextToHit();
            if (view.isSaveChecked())
                view.writeHitCountResults();
            view.updateProgressBar(100);
        }

        public void getTopTermsClickedEvent()
        {
            view.resetProgressBar();
            view.clearDisplayBox("topterm");
            view.clearElapsedTimeLabel("topterm");

            topTermModel = new TopTermModel(view.getHitCountList(), view.getBeginYear(),
view.getEndYear());
            view.setTopTermModel(topTermModel);
            System.DateTime startTime = System.DateTime.Now;
            if (view.getTopTermMethodName().Equals("Early Growth"))
                topTermModel.calculateRanksUsingLogRatio();
            else
                topTermModel.calculateRanksUsingPercentage();
            view.showElapsedTimeinaLabel("topterm", (System.DateTime.Now -
startTime).TotalSeconds.ToString());
            view.updateTopTermListBox();
            if (view.isSaveChecked())
                view.writeTopTermResults();
            view.updateProgressBar(100);
        }
        public void getRefineTermsClickedEvent()
        {
            view.resetProgressBar();
            view.clearElapsedTimeLabel("term");
            view.disableButtonsNextToTerm();
            System.DateTime startTime = System.DateTime.Now;
            termModel =
                new TermCollectionModel(view.getSeedTerm(), view.getTermSourceList(),
view.getRegistry(), view.getMaxThreads(), view.getMaxTermListSize());
            termModel.setGeneratedTerms(view.getGeneratedTerms());
            termModel.registerProgressListener(this);
            view.clearDisplayBox("term");
            view.setTermModel(termModel);
            termModel.refine();
            view.showElapsedTimeinaLabel("term", (System.DateTime.Now -
startTime).TotalSeconds.ToString());
            view.updateTermListBox();
            if (termModel.getGeneratedTerms().Count > 0)
                view.enableButtonsNextToTerm();
            if (view.isSaveChecked())
                view.writeTermListResults();
            view.updateProgressBar(100);


        }
        public void updateMessage(string message)
        {
            view.updateMessage(message);
        }
```

```
    }

}
```

DeskTopApplication View
Form1.cs

```
using System.ComponentModel;
using System.Data;
using System.Drawing;

using System.Text;
using System.Windows.Forms;
using System.Collections;
using System.Collections.Specialized;
using System.IO;
using LongTailApp.Controller;


namespace LongTailApp
{
    public partial class Form1 : Form
    {
        string specFileListFileName = "specFiles.txt";
        string specHitFileListFileName = "specFilesHit.txt";
        LongTailController controller;
        TermCollectionModel termModel;
        private HitCollectionModel hitModel;
        private TopTermModel topTermModel;
        string executablePath = Application.StartupPath;
        string pathSeparator = ""+Path.DirectorySeparatorChar;
        public Form1()
        {
            InitializeComponent();
            controller = new LongTailController(this);
        }

        public void setTermModel(TermCollectionModel model){
           termModel = model;
        }
        public void setTopTermModel(TopTermModel model)
        {
            topTermModel = model;
        }
        public void setHitModel(HitCollectionModel model)
        {
            hitModel = model;
        }
        private string getHitCountFileLocation()
        {
            string fileLocation = txtResultFileLocation.Text;
            if (!fileLocation.EndsWith("\\"))
                fileLocation += "\\";
            return fileLocation + "hitcounts" + DateTime.Now.ToFileTimeUtc() + ".csv";
        }

        public List<string[]> getHitCountList()
        {
            List<string[]> hitCountList = new List<string[]>();
            for (int i = 0; i < lstBoxHitCounts.Items.Count; i++)
            {
                string[] row = lstBoxHitCounts.Items[i].ToString().Split(',');
                hitCountList.Add(row);

            }
            return hitCountList;
        }

        public int getBeginYear()
```

```
{
    int bYear = 0;
    try
    {
        bYear = int.Parse(txtBeginYear.Text);
    }
    catch (Exception ex)
    {
        txtMessageBox.Text = ex.Message;
    }
    return bYear;
}
public int getEndYear()
{
    int eYear = 0;
    try
    {
        eYear = int.Parse(txtEndYear.Text);
    }
    catch (Exception ex)
    {
        txtMessageBox.Text = ex.Message;
    }
    return eYear;
}
public string getTopTermMethodName()
{
    return ddlistAnalysisMethod.SelectedItem.ToString();
}
public void writeHitCountResults()
{
    List<string[]> generatedHitCounts = hitModel.getGeneratedHitCounts();
    if (!chkBoxSave.Checked) return;
    string hitCountListFile = getHitCountFileLocation();
    try
    {
        StreamWriter sw = new StreamWriter(hitCountListFile, false);
        for (int i = 0; i < generatedHitCounts.Count; i++)
        {
            string[] row = generatedHitCounts[i];
            sw.WriteLine(row[0] + "," + row[1] + "," + row[2]);
        }
        sw.Close();
    }
    catch (Exception ex)
    {
        txtMessageBox.Text += ex.Message+"\n";
    }
}

public void updateTopTermListBox()
{
    ArrayList logs = topTermModel.getRanks();
    for (int i = 0; i < logs.Count; i++)
    {
        string[] r = (string[])logs[i];
        lstTopTerms.Items.Add(r[0] + "," + r[1]);
    }

}
public void updateTermListBox()
{
    int max = getMaxTermListSize();
    List<string> generatedTerms = termModel.getGeneratedTerms();

    for (int i = 0; i < generatedTerms.Count && i < max ; i++)
        lstBoxTermList.Items.Add(generatedTerms[i]);
}

public void enableButtonsNextToTerm()
{
```

```
            btnGetHits.Enabled = true;
            btnRefine.Enabled = true;
        }
        public void disableButtonsNextToHit()
        {
            btnGetTop.Enabled = false;

        }
        public void enableButtonsNextToHit()
        {
            btnGetTop.Enabled = true;

        }
        public int getMaxThreads()
        {
            int maxThreads=10;
            try
            {

                maxThreads = int.Parse(txtPOETermNumber.Text);
            }
            catch (Exception ex)
            {
                txtMessageBox.Text = ex.Message;
            }
            return maxThreads;
        }
        public int getMaxTermListSize()
        {
            int maxTermListSize = 80;
            try
            {
                maxTermListSize = int.Parse(txtTermListSize.Text);

            }
            catch (Exception ex)
            {
                txtMessageBox.Text = ex.Message;
            }
            return maxTermListSize;
        }




        private List<int> getYears()
        {
            List<int> years = new List<int>();
            int bYear = 0;
            int eYear = 0;
            years = new List<int>();
            try
            {
                bYear = int.Parse(txtBeginYear.Text);
                eYear = int.Parse(txtEndYear.Text);
                if ((eYear - bYear) < 0)
                    throw new Exception("eYear cannot be less than bYear");
                else if ((eYear - bYear) > 10)
                    throw new Exception("eYear and bYear cannot be separated by more than
10 years for now");
            }
            catch (Exception ex)
            {
                txtMessageBox.Text= ex.Message;
            }
            for (int i = bYear; i < eYear + 1; i++)
            {
                years.Add(i);
            }
            return years;
```

```
}
public List<string[]> getHitCountInput()
{
    List<int> years = getYears();
    List<string[]> hitCountInput = new List<string[]>();
    List<string> generatedTerms = getGeneratedTerms();
    for (int i = 0; i < generatedTerms.Count; i++)
    {


        for (int j = 0; j < years.Count; j++)
        {
            string[] row = new string[2];
            if (chkBoxAddSeedTerm.Checked)
            {
                if (!generatedTerms[i].ToLower().Contains(txtSeedTerm.Text))
                    row[0] = txtSeedTerm.Text + " " + generatedTerms[i];
                else
                    row[0] = generatedTerms[i];
            }
            else
                row[0] = generatedTerms[i];
            row[1] = years[j].ToString();
            hitCountInput.Add(row);
        }
    }
    return hitCountInput;
}
public void updateMessage(string message)
{
    txtMessageBox.Text = message;
}

public  void resetProgressBar()
{
    updateProgressBar(0);
}
public void updateProgressBar(int value)
{
    if (value > 100) value = 100;
    progressBar1.Value=value;
    progressBar1.Update();
    //progressBar1.Increment(value - progressBar1.Value);


}





private string getTopTermFileLocation()
{
    string fileLocation = txtResultFileLocation.Text;
    if (!fileLocation.EndsWith("\\"))
        fileLocation += "\\";
    return fileLocation + "toptermlist" + DateTime.Now.ToFileTimeUtc() + ".csv";
}



public void disableButtonsNextToTerm()
{
    btnGetHits.Enabled = false;
    btnRefine.Enabled = false;
}



public string getRegistry()
{
    string registry = txtRegistry.Text;
    if (!registry.ToLower().StartsWith("http://"))
        registry = executablePath + pathSeparator + registry;
    if (!registry.EndsWith("" + pathSeparator)) registry += pathSeparator;
    return registry;
}

public string getSeedTerm()
{
    return txtSeedTerm.Text;
}
public ArrayList getHitSourceList()
{
    ArrayList sourceList = new ArrayList();
    sourceList.Add(ddlistHitCountSource.SelectedItem.ToString());
    return sourceList;
}
public ArrayList getTermSourceList()
{
    ArrayList sourceList = new ArrayList();
    for (int i = 0; i < chkboxlistTermGenerationSources.CheckedItems.Count; i++)
        sourceList.Add(chkboxlistTermGenerationSources.CheckedItems[i]);
    return sourceList;
}
private string getTermFileLocation()
{
    string fileLocation = txtResultFileLocation.Text;
    if (!fileLocation.EndsWith("\\"))
        fileLocation += "\\";
    return fileLocation + "termlist" + DateTime.Now.ToFileTimeUtc() + ".csv";
}
public void writeTermListResults()
{
    List<string> generatedTerms = getGeneratedTerms();
    if (!chkBoxSave.Checked) return;
    string termListFile = getTermFileLocation();
    try
    {
        StreamWriter sw = new StreamWriter(termListFile, false);
        for (int i = 0; i < generatedTerms.Count; i++)
            sw.WriteLine(generatedTerms[i]);
        sw.Close();
    }
    catch (Exception ex)
    {
        txtMessageBox.Text = ex.Message;
    }
}
public bool isSaveChecked()
{
    return chkBoxSave.Checked;
}
private void btnGetTerms_Click(object sender, EventArgs e)
{
    controller.getTermsClickedEvent();
}

private void btnGetHits_Click(object sender, EventArgs e)
{
    controller.getHitCountsClickedEvent();
}


private void btnGetTop_Click(object sender, EventArgs e)
{
    controller.getTopTermsClickedEvent();
}
public void writeTopTermResults()
{
    ArrayList logs = topTermModel.getRanks();
    if (!chkBoxSave.Checked) return;
```

```
            string termListFile = getTopTermFileLocation();
            try
            {
                StreamWriter sw = new StreamWriter(termListFile, false);
                for (int i = 0; i < logs.Count; i++)
                {
                    string[] r = (string[])logs[i];
                    sw.WriteLine(r[0] + "," + r[1]);
                }
                sw.Close();
            }
            catch (Exception ex)
            {
                txtMessageBox.Text = ex.Message;
            }
        }
        private void btnGetAll_Click(object sender, EventArgs e)
        {
            this.btnGetTerms_Click(null, null);
            this.btnGetHits_Click(null, null);
            this.btnGetTop_Click(null, null);
        }

        private void Form1_Load(object sender, EventArgs e)
        {
            loadSpecFileList();
            loadHitSpecFileList();
            ddlistAnalysisMethod.SelectedItem = "Early Growth";
            ddlistHitCountSource.SelectedItem = "scirushits";
            chkboxlistTermGenerationSources.SetItemChecked(0, true);
        }

        private void btnRefine_Click(object sender, EventArgs e)
        {
            controller.getRefineTermsClickedEvent();
        }

        public void updateHitCountListBox()
        {
            List<string[]> generatedHitCounts=hitModel.getGeneratedHitCounts();
            for (int i = 0; i < generatedHitCounts.Count; i++)
            {
                string[] row = generatedHitCounts[i];
                lstBoxHitCounts.Items.Add(row[0] + "," + row[1].Replace(",", "") + "," +
row[2]);
            }
        }
        public List<string> getGeneratedTerms()
        {
            List<string> generatedTerms = new List<string>();
            for (int i = 0; i < lstBoxTermList.Items.Count; i++)
                generatedTerms.Add(lstBoxTermList.Items[i].ToString());
            return generatedTerms;
        }
        public void clearElapsedTimeLabel(string label)
        {
            if (label.ToLower().Equals("term"))
                clearElapsedTimeLabel(lblTermsTime);
            else if (label.ToLower().Equals("hit"))
                clearElapsedTimeLabel(lblHitCountsTime);
            else if (label.ToLower().Equals("topterm"))
                clearElapsedTimeLabel(lblTopTermsTime);
        }

        public void showElapsedTimeinaLabel(string label, string time)
        {
            if (label.ToLower().Equals("term"))
                showElapsedTimeinaLabel(lblTermsTime, time);
            else if (label.ToLower().Equals("hit"))
                showElapsedTimeinaLabel(lblHitCountsTime, time);
            else if (label.ToLower().Equals("topterm"))
```

```
                showElapsedTimeinaLabel(lblTopTermsTime, time);
        }
        private void showElapsedTimeinaLabel(Label elapsedTimeLabel, string
elapsedTimeInSeconds)
        {
            elapsedTimeLabel.Text = "Elapsed time:" + elapsedTimeInSeconds + " seconds";
        }
        private void clearElapsedTimeLabel(Label elapsedTimeLabel)
        {
            elapsedTimeLabel.Text = "Elapsed time:";
        }
        public void clearDisplayBox(string label)
        {
            if (label.ToLower().Equals("term"))
                clearDisplayBox(lstBoxTermList);
            else if (label.ToLower().Equals("hit"))
                clearDisplayBox(lstBoxHitCounts);
            else if (label.ToLower().Equals("topterm"))
                clearDisplayBox(lstTopTerms);
        }
        private void clearDisplayBox(ListBox displayBox)
        {
            displayBox.Items.Clear();

        }

        private void btnAddNewHitCountSource_Click(object sender, EventArgs e)
        {
            openFileDialogHit.ShowDialog();
        }

        private void btnAddNewTermSource_Click(object sender, EventArgs e)
        {
            openFileDialogTerm.ShowDialog();

        }

        private void openFileDialogTerm_FileOk(object sender, CancelEventArgs e)
        {
            File.Copy(openFileDialogTerm.FileName,
getRegistry()+openFileDialogTerm.SafeFileName,true);

chkboxlistTermGenerationSources.Items.Add(openFileDialogTerm.SafeFileName.Replace(".xml",
""));
            writeSpecFileList();

        }
        private void loadSpecFileList()
        {

            chkboxlistTermGenerationSources.Items.Clear();
            try
            {
                StreamReader sr = new StreamReader(getRegistry() + specFileListFileName);
                do
                {
                    chkboxlistTermGenerationSources.Items.Add(sr.ReadLine());

                } while (sr.Peek() != -1);
                sr.Close();
            }
            catch (Exception ex)
            {
                txtMessageBox.Text = ex.Message;
            }
        }

        private void writeSpecFileList()
        {
            try
            {
                StreamWriter sw = new StreamWriter(getRegistry()+specFileListFileName,
false);
                for (int i = 0; i < chkboxlistTermGenerationSources.Items.Count; i++)
                    sw.WriteLine(chkboxlistTermGenerationSources.Items[i].ToString());
                sw.Close();
            }
            catch (Exception ex)
            {
                txtMessageBox.Text = ex.Message;
            }
        }
        private void loadHitSpecFileList()
        {

            ddlistHitCountSource.Items.Clear();

            try
            {
                StreamReader sr = new StreamReader(getRegistry() +
specHitFileListFileName);
                do
                {
                    ddlistHitCountSource.Items.Add(sr.ReadLine());

                } while (sr.Peek() != -1);
                sr.Close();
            }
            catch (Exception ex)
            {
                txtMessageBox.Text = ex.Message;
            }
        }

        private void writeHitSpecFileList()
        {


            try
            {
                StreamWriter sw = new StreamWriter(getRegistry() +
specHitFileListFileName, false);
                for (int i = 0; i < ddlistHitCountSource.Items.Count; i++)
                    sw.WriteLine(ddlistHitCountSource.Items[i].ToString());
                sw.Close();
            }
            catch (Exception ex)
            {
                txtMessageBox.Text = ex.Message;
            }
        }
        private void openFileDialogHit_FileOk(object sender, CancelEventArgs e)
        {
            File.Copy(openFileDialogHit.FileName, getRegistry() +
openFileDialogHit.SafeFileName, true);
            ddlistHitCountSource.Items.Add(openFileDialogHit.SafeFileName.Replace(".xml",
""));
            writeHitSpecFileList();
        }

        private void btnTermDelete_Click(object sender, EventArgs e)
        {
            for (int i = 0; i < chkboxlistTermGenerationSources.CheckedItems.Count; i++)
            {

chkboxlistTermGenerationSources.Items.Remove(chkboxlistTermGenerationSources.CheckedItems
[i]);
                i--;
            }
            writeSpecFileList();
        }

        private void btnHitDelete_Click(object sender, EventArgs e)
        {
            ddlistHitCountSource.Items.RemoveAt(ddlistHitCountSource.SelectedIndex);
            writeHitSpecFileList();

        }
    }

        }
Form1.Designer.cs

namespace LongTailApp
{
    partial class Form1
    {
        /// <summary>
        /// Required designer variable.
        /// </summary>
        private System.ComponentModel.IContainer components = null;

        /// <summary>
        /// Clean up any resources being used.
        /// </summary>
        /// <param name="disposing">true if managed resources should be disposed;
otherwise, false.</param>
        protected override void Dispose(bool disposing)
        {
            if (disposing && (components != null))
            {
                components.Dispose();
            }
            base.Dispose(disposing);
        }

        #region Windows Form Designer generated code

        /// <summary>
        /// Required method for Designer support - do not modify
        /// the contents of this method with the code editor.
        /// </summary>
        private void InitializeComponent()
        {
            this.label1 = new System.Windows.Forms.Label();
            this.label2 = new System.Windows.Forms.Label();
            this.label3 = new System.Windows.Forms.Label();
            this.label4 = new System.Windows.Forms.Label();
            this.label5 = new System.Windows.Forms.Label();
            this.label6 = new System.Windows.Forms.Label();
            this.label7 = new System.Windows.Forms.Label();
            this.label8 = new System.Windows.Forms.Label();
            this.txtSeedTerm = new System.Windows.Forms.TextBox();
            this.chkboxlistTermGenerationSources = new
System.Windows.Forms.CheckedListBox();
            this.txtTermListSize = new System.Windows.Forms.TextBox();
            this.ddlistHitCountSource = new System.Windows.Forms.ComboBox();
            this.txtBeginYear = new System.Windows.Forms.TextBox();
            this.txtEndYear = new System.Windows.Forms.TextBox();
            this.ddlistAnalysisMethod = new System.Windows.Forms.ComboBox();
            this.txtPOETermNumber = new System.Windows.Forms.TextBox();
            this.txtRegistry = new System.Windows.Forms.TextBox();
            this.lstBoxTermList = new System.Windows.Forms.ListBox();
            this.lstBoxHitCounts = new System.Windows.Forms.ListBox();
            this.lstTopTerms = new System.Windows.Forms.ListBox();
            this.btnGetTerms = new System.Windows.Forms.Button();
            this.btnGetHits = new System.Windows.Forms.Button();
            this.btnGetTop = new System.Windows.Forms.Button();
            this.btnGetAll = new System.Windows.Forms.Button();
            this.lblTermsTime = new System.Windows.Forms.Label();
            this.lblHitCountsTime = new System.Windows.Forms.Label();
this.lblTopTermsTime = new System.Windows.Forms.Label();
this.btnRefine = new System.Windows.Forms.Button();
this.label9 = new System.Windows.Forms.Label();
this.txtResultFileLocation = new System.Windows.Forms.TextBox();
this.txtMessageBox = new System.Windows.Forms.TextBox();
this.label10 = new System.Windows.Forms.Label();
this.chkBoxAddSeedTerm = new System.Windows.Forms.CheckBox();
this.progressBar1 = new System.Windows.Forms.ProgressBar();
this.chkBoxSave = new System.Windows.Forms.CheckBox();
this.btnAddNewHitCountSource = new System.Windows.Forms.Button();
this.btnAddNewTermSource = new System.Windows.Forms.Button();
this.openFileDialogTerm = new System.Windows.Forms.OpenFileDialog();
this.openFileDialogHit = new System.Windows.Forms.OpenFileDialog();
this.btnTermDelete = new System.Windows.Forms.Button();
this.btnHitDelete = new System.Windows.Forms.Button();
this.SuspendLayout();
//
// label1
//
this.label1.AutoSize = true;
this.label1.Location = new System.Drawing.Point(22, 24);
this.label1.Name = "label1";
this.label1.Size = new System.Drawing.Size(62, 13);
this.label1.TabIndex = 0;
this.label1.Text = "Seed Term:";
//
// label2
//
this.label2.AutoSize = true;
this.label2.Location = new System.Drawing.Point(335, 24);
this.label2.Name = "label2";
this.label2.Size = new System.Drawing.Size(131, 13);
this.label2.TabIndex = 1;
this.label2.Text = "Term Generation Sources:";
//
// label3
//
this.label3.AutoSize = true;
this.label3.Location = new System.Drawing.Point(22, 130);
this.label3.Name = "label3";
this.label3.Size = new System.Drawing.Size(137, 13);
this.label3.TabIndex = 2;
this.label3.Text = "Approximate Term List Size:";
//
// label4
//
this.label4.AutoSize = true;
this.label4.Location = new System.Drawing.Point(335, 129);
this.label4.Name = "label4";
this.label4.Size = new System.Drawing.Size(91, 13);
this.label4.TabIndex = 3;
this.label4.Text = "Hit Count Source:";
//
// label5
//
this.label5.AutoSize = true;
this.label5.Location = new System.Drawing.Point(22, 188);
this.label5.Name = "label5";
this.label5.Size = new System.Drawing.Size(89, 13);
this.label5.TabIndex = 4;
this.label5.Text = "Begin-End Years:";
//
// label6
//
this.label6.AutoSize = true;
this.label6.Location = new System.Drawing.Point(335, 190);
this.label6.Name = "label6";
this.label6.Size = new System.Drawing.Size(87, 13);
this.label6.TabIndex = 5;
this.label6.Text = "Analysis Method:";
//
```

```
            // label7
            //
            this.label7.AutoSize = true;
            this.label7.Location = new System.Drawing.Point(22, 246);
            this.label7.Name = "label7";
            this.label7.Size = new System.Drawing.Size(101, 13);
            this.label7.TabIndex = 6;
            this.label7.Text = "Number of Threads:";
            //
            // label8
            //
            this.label8.AutoSize = true;
            this.label8.Location = new System.Drawing.Point(335, 246);
            this.label8.Name = "label8";
            this.label8.Size = new System.Drawing.Size(48, 13);
            this.label8.TabIndex = 7;
            this.label8.Text = "Registry:";
            //
            // txtSeedTerm
            //
            this.txtSeedTerm.Location = new System.Drawing.Point(178, 24);
            this.txtSeedTerm.Name = "txtSeedTerm";
            this.txtSeedTerm.Size = new System.Drawing.Size(100, 20);
            this.txtSeedTerm.TabIndex = 8;
            this.txtSeedTerm.Text = "renewable energy";
            //
            // chkboxlistTermGenerationSources
            //
            this.chkboxlistTermGenerationSources.FormattingEnabled = true;
            this.chkboxlistTermGenerationSources.Location = new System.Drawing.Point(474,
9);
            this.chkboxlistTermGenerationSources.Name =
"chkboxlistTermGenerationSources";
            this.chkboxlistTermGenerationSources.Size = new System.Drawing.Size(237, 94);
            this.chkboxlistTermGenerationSources.TabIndex = 10;
            //
            // txtTermListSize
            //
            this.txtTermListSize.Location = new System.Drawing.Point(178, 128);
            this.txtTermListSize.Name = "txtTermListSize";
            this.txtTermListSize.Size = new System.Drawing.Size(100, 20);
            this.txtTermListSize.TabIndex = 11;
            this.txtTermListSize.Text = "100";
            //
            // ddlistHitCountSource
            //
            this.ddlistHitCountSource.DisplayMember = "scirushits";
            this.ddlistHitCountSource.FormattingEnabled = true;
            this.ddlistHitCountSource.Items.AddRange(new object[] {
            "scirushits",
            "gscholar"});
            this.ddlistHitCountSource.Location = new System.Drawing.Point(474, 130);
            this.ddlistHitCountSource.Name = "ddlistHitCountSource";
            this.ddlistHitCountSource.Size = new System.Drawing.Size(121, 21);
            this.ddlistHitCountSource.TabIndex = 12;
            //
            // txtBeginYear
            //
            this.txtBeginYear.Location = new System.Drawing.Point(178, 184);
            this.txtBeginYear.Name = "txtBeginYear";
            this.txtBeginYear.Size = new System.Drawing.Size(61, 20);
            this.txtBeginYear.TabIndex = 13;
            this.txtBeginYear.Text = "2006";
            //
            // txtEndYear
            //
            this.txtEndYear.Location = new System.Drawing.Point(245, 184);
            this.txtEndYear.Name = "txtEndYear";
            this.txtEndYear.Size = new System.Drawing.Size(61, 20);
            this.txtEndYear.TabIndex = 14;
            this.txtEndYear.Text = "2008";
```

```
//
// ddlistAnalysisMethod
//
this.ddlistAnalysisMethod.FormattingEnabled = true;
this.ddlistAnalysisMethod.Items.AddRange(new object[] {
"Total Growth",
"Early Growth"});
this.ddlistAnalysisMethod.Location = new System.Drawing.Point(474, 183);
this.ddlistAnalysisMethod.Name = "ddlistAnalysisMethod";
this.ddlistAnalysisMethod.Size = new System.Drawing.Size(121, 21);
this.ddlistAnalysisMethod.TabIndex = 15;
//
// txtPOETermNumber
//
this.txtPOETermNumber.Location = new System.Drawing.Point(178, 240);
this.txtPOETermNumber.Name = "txtPOETermNumber";
this.txtPOETermNumber.Size = new System.Drawing.Size(100, 20);
this.txtPOETermNumber.TabIndex = 16;
this.txtPOETermNumber.Text = "10";
//
// txtRegistry
//
this.txtRegistry.Location = new System.Drawing.Point(474, 239);
this.txtRegistry.Name = "txtRegistry";
this.txtRegistry.Size = new System.Drawing.Size(225, 20);
this.txtRegistry.TabIndex = 17;
this.txtRegistry.Text = "Cameleon#";
//
// lstBoxTermList
//
this.lstBoxTermList.FormattingEnabled = true;
this.lstBoxTermList.Location = new System.Drawing.Point(19, 402);
this.lstBoxTermList.Name = "lstBoxTermList";
this.lstBoxTermList.Size = new System.Drawing.Size(214, 160);
this.lstBoxTermList.TabIndex = 18;
//
// lstBoxHitCounts
//
this.lstBoxHitCounts.FormattingEnabled = true;
this.lstBoxHitCounts.Location = new System.Drawing.Point(252, 402);
this.lstBoxHitCounts.Name = "lstBoxHitCounts";
this.lstBoxHitCounts.Size = new System.Drawing.Size(214, 160);
this.lstBoxHitCounts.TabIndex = 19;
//
// lstTopTerms
//
this.lstTopTerms.FormattingEnabled = true;
this.lstTopTerms.Location = new System.Drawing.Point(485, 402);
this.lstTopTerms.Name = "lstTopTerms";
this.lstTopTerms.Size = new System.Drawing.Size(214, 160);
this.lstTopTerms.TabIndex = 20;
//
// btnGetTerms
//
this.btnGetTerms.Location = new System.Drawing.Point(41, 370);
this.btnGetTerms.Name = "btnGetTerms";
this.btnGetTerms.Size = new System.Drawing.Size(75, 23);
this.btnGetTerms.TabIndex = 21;
this.btnGetTerms.Text = "Get Terms";
this.btnGetTerms.UseVisualStyleBackColor = true;
this.btnGetTerms.Click += new System.EventHandler(this.btnGetTerms_Click);
//
// btnGetHits
//
this.btnGetHits.Enabled = false;
this.btnGetHits.Location = new System.Drawing.Point(270, 370);
this.btnGetHits.Name = "btnGetHits";
this.btnGetHits.Size = new System.Drawing.Size(85, 23);
this.btnGetHits.TabIndex = 22;
this.btnGetHits.Text = "Get Hitcounts";
this.btnGetHits.UseVisualStyleBackColor = true;
this.btnGetHits.Click += new System.EventHandler(this.btnGetHits_Click);
//
// btnGetTop
//
this.btnGetTop.Enabled = false;
this.btnGetTop.Location = new System.Drawing.Point(547, 370);
this.btnGetTop.Name = "btnGetTop";
this.btnGetTop.Size = new System.Drawing.Size(89, 23);
this.btnGetTop.TabIndex = 23;
this.btnGetTop.Text = "Get Top Terms";
this.btnGetTop.UseVisualStyleBackColor = true;
this.btnGetTop.Click += new System.EventHandler(this.btnGetTop_Click);
//
// btnGetAll
//
this.btnGetAll.Location = new System.Drawing.Point(322, 341);
this.btnGetAll.Name = "btnGetAll";
this.btnGetAll.Size = new System.Drawing.Size(85, 23);
this.btnGetAll.TabIndex = 24;
this.btnGetAll.Text = "Get All";
this.btnGetAll.UseVisualStyleBackColor = true;
this.btnGetAll.Click += new System.EventHandler(this.btnGetAll_Click);
//
// lblTermsTime
//
this.lblTermsTime.AutoSize = true;
this.lblTermsTime.Location = new System.Drawing.Point(19, 569);
this.lblTermsTime.Name = "lblTermsTime";
this.lblTermsTime.Size = new System.Drawing.Size(70, 13);
this.lblTermsTime.TabIndex = 25;
this.lblTermsTime.Text = "Elapsed time:";
//
// lblHitCountsTime
//
this.lblHitCountsTime.AutoSize = true;
this.lblHitCountsTime.Location = new System.Drawing.Point(249, 572);
this.lblHitCountsTime.Name = "lblHitCountsTime";
this.lblHitCountsTime.Size = new System.Drawing.Size(70, 13);
this.lblHitCountsTime.TabIndex = 26;
this.lblHitCountsTime.Text = "Elapsed time:";
//
// lblTopTermsTime
//
this.lblTopTermsTime.AutoSize = true;
this.lblTopTermsTime.Location = new System.Drawing.Point(482, 572);
this.lblTopTermsTime.Name = "lblTopTermsTime";
this.lblTopTermsTime.Size = new System.Drawing.Size(70, 13);
this.lblTopTermsTime.TabIndex = 27;
this.lblTopTermsTime.Text = "Elapsed time:";
//
// btnRefine
//
this.btnRefine.Enabled = false;
this.btnRefine.Location = new System.Drawing.Point(137, 370);
this.btnRefine.Name = "btnRefine";
this.btnRefine.Size = new System.Drawing.Size(75, 23);
this.btnRefine.TabIndex = 28;
this.btnRefine.Text = "Refine";
this.btnRefine.UseVisualStyleBackColor = true;
this.btnRefine.Click += new System.EventHandler(this.btnRefine_Click);
//
// label9
//
this.label9.AutoSize = true;
this.label9.Location = new System.Drawing.Point(25, 294);
this.label9.Name = "label9";
this.label9.Size = new System.Drawing.Size(138, 13);
this.label9.TabIndex = 29;
this.label9.Text = "Location to store result files:";
//
// txtResultFileLocation
```

```
            //
            this.txtResultFileLocation.Location = new System.Drawing.Point(178, 286);
            this.txtResultFileLocation.Name = "txtResultFileLocation";
            this.txtResultFileLocation.Size = new System.Drawing.Size(100, 20);
            this.txtResultFileLocation.TabIndex = 30;
            this.txtResultFileLocation.Text = "Results";
            //
            // txtMessageBox
            //
            this.txtMessageBox.Location = new System.Drawing.Point(474, 286);
            this.txtMessageBox.Multiline = true;
            this.txtMessageBox.Name = "txtMessageBox";
            this.txtMessageBox.Size = new System.Drawing.Size(225, 46);
            this.txtMessageBox.TabIndex = 31;
            //
            // label10
            //
            this.label10.AutoSize = true;
            this.label10.Location = new System.Drawing.Point(335, 286);
            this.label10.Name = "label10";
            this.label10.Size = new System.Drawing.Size(58, 13);
            this.label10.TabIndex = 32;
            this.label10.Text = "Messages:";
            //
            // chkBoxAddSeedTerm
            //
            this.chkBoxAddSeedTerm.AutoSize = true;
            this.chkBoxAddSeedTerm.Location = new System.Drawing.Point(370, 374);
            this.chkBoxAddSeedTerm.Name = "chkBoxAddSeedTerm";
            this.chkBoxAddSeedTerm.Size = new System.Drawing.Size(93, 17);
            this.chkBoxAddSeedTerm.TabIndex = 33;
            this.chkBoxAddSeedTerm.Text = "add seed term";
            this.chkBoxAddSeedTerm.UseVisualStyleBackColor = true;
            //
            // progressBar1
            //
            this.progressBar1.Location = new System.Drawing.Point(0, 600);
            this.progressBar1.Name = "progressBar1";
            this.progressBar1.Size = new System.Drawing.Size(719, 23);
            this.progressBar1.TabIndex = 34;
            //
            // chkBoxSave
            //
            this.chkBoxSave.AutoSize = true;
            this.chkBoxSave.Location = new System.Drawing.Point(178, 326);
            this.chkBoxSave.Name = "chkBoxSave";
            this.chkBoxSave.Size = new System.Drawing.Size(89, 17);
            this.chkBoxSave.TabIndex = 36;
            this.chkBoxSave.Text = "Save Results";
            this.chkBoxSave.UseVisualStyleBackColor = true;
            //
            // btnAddNewHitCountSource
            //
            this.btnAddNewHitCountSource.Location = new System.Drawing.Point(338, 145);
            this.btnAddNewHitCountSource.Name = "btnAddNewHitCountSource";
            this.btnAddNewHitCountSource.Size = new System.Drawing.Size(75, 23);
            this.btnAddNewHitCountSource.TabIndex = 37;
            this.btnAddNewHitCountSource.Text = "Add New...";
            this.btnAddNewHitCountSource.UseVisualStyleBackColor = true;
            this.btnAddNewHitCountSource.Click += new
System.EventHandler(this.btnAddNewHitCountSource_Click);
            //
            // btnAddNewTermSource
            //
            this.btnAddNewTermSource.Location = new System.Drawing.Point(338, 41);
            this.btnAddNewTermSource.Name = "btnAddNewTermSource";
            this.btnAddNewTermSource.Size = new System.Drawing.Size(75, 23);
            this.btnAddNewTermSource.TabIndex = 38;
            this.btnAddNewTermSource.Text = "Add New...";
            this.btnAddNewTermSource.UseVisualStyleBackColor = true;
            this.btnAddNewTermSource.Click += new
System.EventHandler(this.btnAddNewTermSource_Click);
            //
            // openFileDialogTerm
            //
            this.openFileDialogTerm.FileName = "openFileDialog1";
            this.openFileDialogTerm.FileOk += new
System.ComponentModel.CancelEventHandler(this.openFileDialogTerm_FileOk);
            //
            // openFileDialogHit
            //
            this.openFileDialogHit.FileName = "openFileDialog2";
            this.openFileDialogHit.FileOk += new
System.ComponentModel.CancelEventHandler(this.openFileDialogHit_FileOk);
            //
            // btnTermDelete
            //
            this.btnTermDelete.Location = new System.Drawing.Point(419, 40);
            this.btnTermDelete.Name = "btnTermDelete";
            this.btnTermDelete.Size = new System.Drawing.Size(26, 23);
            this.btnTermDelete.TabIndex = 39;
            this.btnTermDelete.Text = "X";
            this.btnTermDelete.UseVisualStyleBackColor = true;
            this.btnTermDelete.Click += new
System.EventHandler(this.btnTermDelete_Click);
            //
            // btnHitDelete
            //
            this.btnHitDelete.Location = new System.Drawing.Point(419, 145);
            this.btnHitDelete.Name = "btnHitDelete";
            this.btnHitDelete.Size = new System.Drawing.Size(26, 23);
            this.btnHitDelete.TabIndex = 40;
            this.btnHitDelete.Text = "X";
            this.btnHitDelete.UseVisualStyleBackColor = true;
            this.btnHitDelete.Click += new System.EventHandler(this.btnHitDelete_Click);
            //
            // Form1
            //
            this.AutoScaleDimensions = new System.Drawing.SizeF(6F, 13F);
            this.AutoScaleMode = System.Windows.Forms.AutoScaleMode.Font;
            this.ClientSize = new System.Drawing.Size(718, 623);
            this.Controls.Add(this.btnHitDelete);
            this.Controls.Add(this.btnTermDelete);
            this.Controls.Add(this.btnAddNewTermSource);
            this.Controls.Add(this.btnAddNewHitCountSource);
            this.Controls.Add(this.chkBoxSave);
            this.Controls.Add(this.progressBar1);
            this.Controls.Add(this.chkBoxAddSeedTerm);
            this.Controls.Add(this.label10);
            this.Controls.Add(this.txtMessageBox);
            this.Controls.Add(this.txtResultFileLocation);
            this.Controls.Add(this.label9);
            this.Controls.Add(this.btnRefine);
            this.Controls.Add(this.lblTopTermsTime);
            this.Controls.Add(this.lblHitCountsTime);
            this.Controls.Add(this.lblTermsTime);
            this.Controls.Add(this.btnGetAll);
            this.Controls.Add(this.btnGetTop);
            this.Controls.Add(this.btnGetHits);
            this.Controls.Add(this.btnGetTerms);
            this.Controls.Add(this.lstTopTerms);
            this.Controls.Add(this.lstBoxHitCounts);
            this.Controls.Add(this.lstBoxTermList);
            this.Controls.Add(this.txtRegistry);
            this.Controls.Add(this.txtPOETermNumber);
            this.Controls.Add(this.ddlistAnalysisMethod);
            this.Controls.Add(this.txtEndYear);
            this.Controls.Add(this.txtBeginYear);
            this.Controls.Add(this.ddlistHitCountSource);
            this.Controls.Add(this.txtTermListSize);
            this.Controls.Add(this.chkboxlistTermGenerationSources);
            this.Controls.Add(this.txtSeedTerm);
            this.Controls.Add(this.label8);
            this.Controls.Add(this.label7);
            this.Controls.Add(this.label6);
            this.Controls.Add(this.label5);
            this.Controls.Add(this.label4);
            this.Controls.Add(this.label3);
            this.Controls.Add(this.label2);
            this.Controls.Add(this.label1);
            this.Name = "Form1";
            this.Text = "LongTail";
            this.Load += new System.EventHandler(this.Form1_Load);
            this.ResumeLayout(false);
            this.PerformLayout();

        }

        #endregion

        private System.Windows.Forms.Label label1;
        private System.Windows.Forms.Label label2;
        private System.Windows.Forms.Label label3;
        private System.Windows.Forms.Label label4;
        private System.Windows.Forms.Label label5;
        private System.Windows.Forms.Label label6;
        private System.Windows.Forms.Label label7;
        private System.Windows.Forms.Label label8;
        private System.Windows.Forms.TextBox txtSeedTerm;
        private System.Windows.Forms.CheckedListBox chkboxlistTermGenerationSources;
        private System.Windows.Forms.TextBox txtTermListSize;
        private System.Windows.Forms.ComboBox ddlistHitCountSource;
        private System.Windows.Forms.TextBox txtBeginYear;
        private System.Windows.Forms.TextBox txtEndYear;
        private System.Windows.Forms.ComboBox ddlistAnalysisMethod;
        private System.Windows.Forms.TextBox txtPOETermNumber;
        private System.Windows.Forms.TextBox txtRegistry;
        private System.Windows.Forms.ListBox lstBoxTermList;
        private System.Windows.Forms.ListBox lstBoxHitCounts;
        private System.Windows.Forms.ListBox lstTopTerms;
        private System.Windows.Forms.Button btnGetTerms;
        private System.Windows.Forms.Button btnGetHits;
        private System.Windows.Forms.Button btnGetTop;
        private System.Windows.Forms.Button btnGetAll;
        private System.Windows.Forms.Label lblTermsTime;
        private System.Windows.Forms.Label lblHitCountsTime;
        private System.Windows.Forms.Label lblTopTermsTime;
        private System.Windows.Forms.Button btnRefine;
        private System.Windows.Forms.Label label9;
        private System.Windows.Forms.TextBox txtResultFileLocation;
        private System.Windows.Forms.Label label10;
        private System.Windows.Forms.CheckBox chkBoxAddSeedTerm;
        private System.Windows.Forms.ProgressBar progressBar1;
        public System.Windows.Forms.TextBox txtMessageBox;
        private System.Windows.Forms.CheckBox chkBoxSave;
        private System.Windows.Forms.Button btnAddNewHitCountSource;
        private System.Windows.Forms.Button btnAddNewTermSource;
        private System.Windows.Forms.OpenFileDialog openFileDialogTerm;
        private System.Windows.Forms.OpenFileDialog openFileDialogHit;
        private System.Windows.Forms.Button btnTermDelete;
        private System.Windows.Forms.Button btnHitDelete;
    }
}
```

Program.cs

```
using System;
using System.Collections.Generic;

using System.Windows.Forms;

namespace LongTailApp
{
    static class Program
    {
        /// <summary>
        /// The main entry point for the application.
        /// </summary>
        [STAThread]
        static void Main()
        {
            Application.EnableVisualStyles();
            Application.SetCompatibleTextRenderingDefault(false);
            Application.Run(new Form1());
        }
    }
}
```

Web Application View & Controller

```
using System;
using System.Collections.Generic;
using System.Linq;
using System.Web;
using System.Web.UI;
using System.Web.UI.WebControls;
using System.Data;
using System.Configuration;
using System.Web.Security;
using System.Web.UI.WebControls.WebParts;
using System.Web.UI.HtmlControls;
using System.Xml;
using System.Xml.XPath;
using System.Text;
using System.IO;
using System.Collections;
using System.Net;
using System.Threading;

    public partial class MainView : System.Web.UI.Page
    {


        string specFileListFileName = "specFiles.txt";
        string specHitFileListFileName = "specFilesHit.txt";
        LongTailController controller;
        TermCollectionModel termModel;
        private HitCollectionModel hitModel;
        private TopTermModel topTermModel;

        string pathSeparator = "" + Path.DirectorySeparatorChar;




        public void setTermModel(TermCollectionModel model)
        {
            termModel = model;
        }
        public void setTopTermModel(TopTermModel model)
        {
            topTermModel = model;
        }
        public void setHitModel(HitCollectionModel model)
        {
            hitModel = model;
        }
        private string getHitCountFileLocation()
        {
            string fName = "hitcounts" + DateTime.Now.ToFileTimeUtc() + ".csv";
            HyperLink2.NavigateUrl = "~/ResultFiles/" + fName;
            return Server.MapPath("~\\ResultFiles\\" + fName);
```

```
}

public List<string[]> getHitCountList()
{
    List<string[]> hitCountList = new List<string[]>();
    for (int i = 0; i < lstBoxHitCounts.Items.Count; i++)
    {
        string[] row = lstBoxHitCounts.Items[i].Value.Split(',');
        hitCountList.Add(row);

    }
    return hitCountList;
}

public int getBeginYear()
{
    int bYear = 0;
    try
    {
        bYear = int.Parse(txtBeginYear.Text);
    }
    catch (Exception ex)
    {
        txtMessageBox.Text += ex.Message+"\n";
    }
    return bYear;
}
public int getEndYear()
{
    int eYear = 0;
    try
    {
        eYear = int.Parse(txtEndYear.Text);
    }
    catch (Exception ex)
    {
        txtMessageBox.Text += ex.Message+"\n";
    }
    return eYear;
}
public string getTopTermMethodName()
{
    return ddlistAnalysisMethod.SelectedItem.ToString();
}
private List<string[]> getGeneratedHitCounts()
{
    if (hitModel != null)
        return hitModel.getGeneratedHitCounts();
    else
    {
        List<string[]> hitCounts = new List<string[]>();
        for (int i = 0; i < lstBoxHitCounts.Items.Count; i++)
        {
            string[] row = lstBoxHitCounts.Items[i].Value.Split(',');
            hitCounts.Add(row);

        }
        return hitCounts;
    }
}

public void writeHitCountResults()
{
    if (!chkBoxSave.Checked) return;
    List<string[]> generatedHitCounts = getGeneratedHitCounts();

    string hitCountListFile = getHitCountFileLocation();
    try
    {
        StreamWriter sw = new StreamWriter(hitCountListFile, false);
        for (int i = 0; i < generatedHitCounts.Count; i++)
        {
            string[] row = generatedHitCounts[i];
            sw.WriteLine(row[0] + "," + row[1] + "," + row[2]);
        }
        sw.Close();
        HyperLink2.Visible = true;
    }
    catch (Exception ex)
    {
        txtMessageBox.Text += ex.Message + "\n";
    }
}

public void updateTopTermListBox()
{
    ArrayList logs = topTermModel.getRanks();
    for (int i = 0; i < logs.Count; i++)
    {
        string[] r = (string[])logs[i];
        lstTopTerms.Items.Add(r[0] + "," + r[1]);
    }

}
public void updateTermListBox()
{
    int max = getMaxTermListSize();
    List<string> generatedTerms = termModel.getGeneratedTerms();

    for (int i = 0; i < generatedTerms.Count && i < max; i++)
        lstBoxTermList.Items.Add(generatedTerms[i]);
}

public void enableButtonsNextToTerm()
{
    btnGetHits.Enabled = true;
    btnRefine.Enabled = true;
}
public void disableButtonsNextToHit()
{
    btnGetTop.Enabled = false;

}
public void enableButtonsNextToHit()
{
    btnGetTop.Enabled = true;

}
public int getMaxThreads()
{
    int maxThreads = 10;
    try
    {

        maxThreads = int.Parse(txtPOETermNumber.Text);
    }
    catch (Exception ex)
    {
        txtMessageBox.Text += ex.Message+ "\n";
    }
    return maxThreads;
}
public int getMaxTermListSize()
{
    int maxTermListSize = 80;
    try
    {
        maxTermListSize = int.Parse(txtTermListSize.Text);

    }
```

```
            catch (Exception ex)
            {
                txtMessageBox.Text += ex.Message+"\n";
            }
            return maxTermListSize;
        }




        private List<int> getYears()
        {
            List<int> years = new List<int>();
            int bYear = 0;
            int eYear = 0;
            years = new List<int>();
            try
            {
                bYear = int.Parse(txtBeginYear.Text);
                eYear = int.Parse(txtEndYear.Text);
                if ((eYear - bYear) < 0)
                    throw new Exception("eYear cannot be less than bYear");
                else if ((eYear - bYear) > 10)
                    throw new Exception("eYear and bYear cannot be separated by more than
10 years for now");
            }
            catch (Exception ex)
            {
                txtMessageBox.Text += ex.Message+ "\n";
            }
            for (int i = bYear; i < eYear + 1; i++)
            {
                years.Add(i);
            }
            return years;
        }
        public List<string[]> getHitCountInput()
        {
            List<int> years = getYears();
            List<string[]> hitCountInput = new List<string[]>();
            List<string> generatedTerms = getGeneratedTerms();
            for (int i = 0; i < generatedTerms.Count; i++)
            {


                for (int j = 0; j < years.Count; j++)
                {
                    string[] row = new string[2];
                    if (chkBoxAddSeedTerm.Checked)
                    {
                        if (!generatedTerms[i].ToLower().Contains(txtSeedTerm.Text))
                            row[0] = txtSeedTerm.Text + " " + generatedTerms[i];
                        else
                            row[0] = generatedTerms[i];
                    }
                    else
                        row[0] = generatedTerms[i];
                    row[1] = years[j].ToString();
                    hitCountInput.Add(row);
                }
            }
            return hitCountInput;
        }
        public void updateMessage(string message)
        {
            txtMessageBox.Text += message + "\n";
        }


        public List<string> getGeneratedTerms()
```

```
{
    List<string> generatedTerms = new List<string>();
    for (int i = 0; i < lstBoxTermList.Items.Count; i++)
        generatedTerms.Add(lstBoxTermList.Items[i].Value);
    return generatedTerms;
}



private string getTopTermFileLocation()
{
    string fName = "toptermlist" + DateTime.Now.ToFileTimeUtc() + ".csv";
    HyperLink3.NavigateUrl = "~/ResultFiles/" + fName;
    return Server.MapPath("~\\ResultFiles\\" + fName);
}


public void disableButtonsNextToTerm()
{
    btnGetHits.Enabled = false;
    btnRefine.Enabled = false;
}


public string getRegistry()
{
    string registry = txtRegistry.Text;
    if (!registry.ToLower().StartsWith("http://"))
        registry = Server.MapPath("~/") + registry;
    if (!registry.EndsWith("" + pathSeparator)) registry += pathSeparator;
    return registry;
}

public string getSeedTerm()
{
    return txtSeedTerm.Text;
}
public ArrayList getHitSourceList()
{
    ArrayList sourceList = new ArrayList();
    sourceList.Add(ddlistHitCountSource.SelectedItem.Value);
    return sourceList;
}
public ArrayList getTermSourceList()
{
    ArrayList sourceList = new ArrayList();
    for (int i = 0; i < chkboxlistTermGenerationSources.Items.Count; i++)
        if (chkboxlistTermGenerationSources.Items[i].Selected)
            sourceList.Add(chkboxlistTermGenerationSources.Items[i].Value);
    return sourceList;
}
private string getTermFileLocation()
{
    string fName = "termlist"+DateTime.Now.ToFileTimeUtc() + ".csv";
    HyperLink1.NavigateUrl = "~/ResultFiles/"+fName;
    return Server.MapPath("~\\ResultFiles\\"+fName);
}
public void writeTermListResults()
{
    if (!chkBoxSave.Checked) return;
    List<string> generatedTerms = getGeneratedTerms();

    string termListFile = getTermFileLocation();
    try
    {
        StreamWriter sw = new StreamWriter(termListFile, false);
        for (int i = 0; i < generatedTerms.Count; i++)
```

```
            sw.WriteLine(generatedTerms[i]);
        sw.Close();
        HyperLink1.Visible = true;
    }
    catch (Exception ex)
    {
        txtMessageBox.Text += ex.Message +"\n";
    }
}
public bool isSaveChecked()
{
    return chkBoxSave.Checked;
}

private ArrayList getLogs()
{
    if (topTermModel != null)
        return topTermModel.getRanks();
    else
    {
        ArrayList logs = new ArrayList();
        for (int i = 0; i < lstTopTerms.Items.Count; i++)
        {
            string[] row = lstTopTerms.Items[i].Value.Split(',');
            logs.Add(row.Clone());
        }
        return logs;
    }
}

public void writeTopTermResults()
{
    ArrayList logs = getLogs();
    if (!chkBoxSave.Checked) return;
    string termListFile = getTopTermFileLocation();
    try
    {
        StreamWriter sw = new StreamWriter(termListFile, false);
        for (int i = 0; i < logs.Count; i++)
        {
            string[] r = (string[])logs[i];
            sw.WriteLine(r[0] + "," + r[1]);
        }
        sw.Close();
        HyperLink3.Visible = true;
    }
    catch (Exception ex)
    {
        txtMessageBox.Text += ex.Message + "\n";
    }
}
private void saveFiles()
{
    if (lstBoxHitCounts.Items.Count>0)
    this.writeHitCountResults();
    if (lstBoxTermList.Items.Count>0)
    this.writeTermListResults();
    if (lstTopTerms.Items.Count>0)
    this.writeTopTermResults();
}
private void initialLoad()
{

    loadSpecFileList();
    loadHitSpecFileList();
    ddlistAnalysisMethod.SelectedValue = "Early Growth";
    ddlistHitCountSource.SelectedValue = "scirushits";
    chkboxlistTermGenerationSources.SelectedIndex = 0;
}

protected void btnRefine_Click(object sender, EventArgs e)
```

```
        {
            controller.getRefineTermsClickedEvent();
        }

        public void updateHitCountListBox()
        {
            List<string[]> generatedHitCounts = hitModel.getGeneratedHitCounts();
            for (int i = 0; i < generatedHitCounts.Count; i++)
            {
                string[] row = generatedHitCounts[i];
                lstBoxHitCounts.Items.Add(row[0] + "," + row[1].Replace(",", "") + "," +
row[2]);
            }
        }

        public void clearElapsedTimeLabel(string label)
        {
            if (label.ToLower().Equals("term"))
                clearElapsedTimeLabel(lblTermsTime);
            else if (label.ToLower().Equals("hit"))
                clearElapsedTimeLabel(lblHitCountsTime);
            else if (label.ToLower().Equals("topterm"))
                clearElapsedTimeLabel(lblTopTermsTime);
        }

        public void showElapsedTimeinaLabel(string label, string time)
        {
            if (label.ToLower().Equals("term"))
                showElapsedTimeinaLabel(lblTermsTime, time);
            else if (label.ToLower().Equals("hit"))
                showElapsedTimeinaLabel(lblHitCountsTime, time);
            else if (label.ToLower().Equals("topterm"))
                showElapsedTimeinaLabel(lblTopTermsTime, time);
        }
        private void showElapsedTimeinaLabel(Label elapsedTimeLabel, string
elapsedTimeInSeconds)
        {
            elapsedTimeLabel.Text = "Elapsed time:" + elapsedTimeInSeconds + " seconds";
        }
        private void clearElapsedTimeLabel(Label elapsedTimeLabel)
        {
            elapsedTimeLabel.Text = "Elapsed time:";
        }
        public void clearDisplayBox(string label)
        {
            if (label.ToLower().Equals("term"))
                clearDisplayBox(lstBoxTermList);
            else if (label.ToLower().Equals("hit"))
                clearDisplayBox(lstBoxHitCounts);
            else if (label.ToLower().Equals("topterm"))
                clearDisplayBox(lstTopTerms);
        }
        private void clearDisplayBox(ListBox displayBox)
        {
            displayBox.Items.Clear();

        }



        private void loadSpecFileList()
        {

            chkboxlistTermGenerationSources.Items.Clear();
            try
            {
                StreamReader sr = new StreamReader(Server.MapPath("~/") +
specFileListFileName);
                do
                {
                    chkboxlistTermGenerationSources.Items.Add(sr.ReadLine());

                } while (sr.Peek() != -1);
                sr.Close();
            }
            catch (Exception ex)
            {
                txtMessageBox.Text += ex.Message +"\n";
            }
        }

        private void writeSpecFileList()
        {


            try
            {
                StreamWriter sw = new StreamWriter(getRegistry() + specFileListFileName,
false);
                for (int i = 0; i < chkboxlistTermGenerationSources.Items.Count; i++)
                    sw.WriteLine(chkboxlistTermGenerationSources.Items[i].Value);
                sw.Close();
            }
            catch (Exception ex)
            {
                txtMessageBox.Text += ex.Message + "\n";
            }
        }
        private void loadHitSpecFileList()
        {

            ddlistHitCountSource.Items.Clear();

            try
            {
                StreamReader sr = new
StreamReader(Server.MapPath("~/")+specHitFileListFileName);
                do
                {
                    ddlistHitCountSource.Items.Add(sr.ReadLine());

                } while (sr.Peek() != -1);
                sr.Close();
            }
            catch (Exception ex)
            {
                txtMessageBox.Text += ex.Message +"\n";
            }
        }

        private void writeHitSpecFileList()
        {


            try
            {
                StreamWriter sw = new StreamWriter(getRegistry() +
specHitFileListFileName, false);
                for (int i = 0; i < ddlistHitCountSource.Items.Count; i++)
                    sw.WriteLine(ddlistHitCountSource.Items[i].Value);
                sw.Close();
            }
            catch (Exception ex)
            {
                txtMessageBox.Text += ex.Message+"\n";
            }
        }
        private void btnHitDelete_Click(object sender, EventArgs e)
        {


        }


        protected void btnAnalyze_Click(object sender, EventArgs e)
        {
            controller.getTermsClickedEvent();
        }





        protected void btnHitCounts_Click(object sender, EventArgs e)
        {
            controller.getHitCountsClickedEvent();
        }



        protected void Button1_Click(object sender, EventArgs e)
        {
            controller.getTopTermsClickedEvent();
        }




        protected void Page_Load(object sender, EventArgs e)
        {
            if (!Page.IsPostBack) initialLoad();
            controller = new LongTailController(this);

        }
        protected void Button2_Click(object sender, EventArgs e)
        {
            btnAnalyze_Click(null, null);
            btnHitCounts_Click(null, null);
            Button1_Click(null, null);
        }



        public void setMessageText(string message)
        {
            txtMessageBox.Text += message+"\n";
        }




        protected void Button3_Click(object sender, EventArgs e)
        {
            chkboxlistTermGenerationSources.Items.Add(new
ListItem(txtNewTermSource.Text));

        }
        protected void btnAddNewHitSource_Click(object sender, EventArgs e)
        {
            ddlistHitCountSource.Items.Add(new ListItem(txtNewHitSource.Text));
        }
        protected void btnDeleteTermSource_Click(object sender, EventArgs e)
        {
            for (int i = 0; i < chkboxlistTermGenerationSources.Items.Count; i++)
            {
                if (chkboxlistTermGenerationSources.Items[i].Selected)
                    chkboxlistTermGenerationSources.Items.RemoveAt(i);
                i--;
            }

        }
        protected void btnDeleteHitSource_Click(object sender, EventArgs e)
        {
            ddlistHitCountSource.Items.RemoveAt(ddlistHitCountSource.SelectedIndex);

        }
        protected void chkBoxSave_CheckedChanged(object sender, EventArgs e)
        {
            if (chkBoxSave.Checked)
            {
                saveFiles();
            }
        }

}
    public class LongTailController : ErrorListener
    {
        protected MainView view;
        TopTermModel topTermModel;
        HitCollectionModel hitCollectionModel;
        TermCollectionModel termModel;

        public void updateError(string message)
        {
            view.setMessageText(message);
        }

        public LongTailController(MainView appForm)
        {
            view = appForm;

        }

        public void getTermsClickedEvent()
        {

            view.clearDisplayBox("term");
            view.clearElapsedTimeLabel("term");
            view.disableButtonsNextToTerm();
            termModel =
                new TermCollectionModel(view.getSeedTerm(), view.getTermSourceList(),
view.getRegistry(), view.getMaxThreads(), view.getMaxTermListSize());
            termModel.registerErrorListener(this);
            System.DateTime startTime = System.DateTime.Now;

            view.setTermModel(termModel);
            termModel.collectTerms();
            view.showElapsedTimeinaLabel("term", (System.DateTime.Now -
startTime).TotalSeconds.ToString());
            view.updateTermListBox();
            view.updateMessage("Total of " +
Math.Min(termModel.getGeneratedTerms().Count, view.getMaxTermListSize()) + " terms");
            if (termModel.getGeneratedTerms().Count > 0)
                view.enableButtonsNextToTerm();
            if (view.isSaveChecked())
                view.writeTermListResults();

        }
        public void getHitCountsClickedEvent()
        {
            view.clearDisplayBox("hit");
            view.clearElapsedTimeLabel("hit");
            view.disableButtonsNextToHit();
            hitCollectionModel =
                new HitCollectionModel(view.getSeedTerm(), view.getHitSourceList(),
view.getRegistry(), view.getMaxThreads(), view.getMaxTermListSize(),
view.getHitCountInput());
            hitCollectionModel.registerErrorListener(this);
            System.DateTime startTime = System.DateTime.Now;
            view.setHitModel(hitCollectionModel);
            hitCollectionModel.collectResults();
            view.showElapsedTimeinaLabel("hit", (System.DateTime.Now -
startTime).TotalSeconds.ToString());
            view.updateHitCountListBox();
            if (hitCollectionModel.getGeneratedHitCounts().Count > 1)
                view.enableButtonsNextToHit();
            if (view.isSaveChecked())
                view.writeHitCountResults();
        }

        public void getTopTermsClickedEvent()
        {
            view.clearDisplayBox("topterm");
            view.clearElapsedTimeLabel("topterm");

            topTermModel = new TopTermModel(view.getHitCountList(), view.getBeginYear(),
view.getEndYear());
            view.setTopTermModel(topTermModel);
            System.DateTime startTime = System.DateTime.Now;
            if (view.getTopTermMethodName().Equals("Early Growth"))
                topTermModel.calculateRanksUsingLogRatio();
            else
                topTermModel.calculateRanksUsingPercentage();
            view.showElapsedTimeinaLabel("topterm", (System.DateTime.Now -
startTime).TotalSeconds.ToString());
            view.updateTopTermListBox();
            if (view.isSaveChecked())
                view.writeTopTermResults();
        }
        public void getRefineTermsClickedEvent()
        {

            view.clearElapsedTimeLabel("term");
            view.disableButtonsNextToTerm();
            System.DateTime startTime = System.DateTime.Now;
            termModel =
                new TermCollectionModel(view.getSeedTerm(), view.getTermSourceList(),
view.getRegistry(), view.getMaxThreads(), view.getMaxTermListSize());
            termModel.setGeneratedTerms(view.getGeneratedTerms());
            view.clearDisplayBox("term");
            view.setTermModel(termModel);
            termModel.refine();
            view.showElapsedTimeinaLabel("term", (System.DateTime.Now -
startTime).TotalSeconds.ToString());
            view.updateTermListBox();
            if (termModel.getGeneratedTerms().Count > 0)
                view.enableButtonsNextToTerm();
            if (view.isSaveChecked())
                view.writeTermListResults();

        }
        public void updateMessage(string message)
        {
            view.updateMessage(message);
        }



    }

<%@ Page Language="C#" AutoEventWireup="true"  CodeFile="Default.aspx.cs"
Inherits="MainView" %>

<!DOCTYPE html PUBLIC "-//W3C//DTD XHTML 1.0 Transitional//EN"
"http://www.w3.org/TR/xhtml1/DTD/xhtml1-transitional.dtd">

<html xmlns="http://www.w3.org/1999/xhtml">
<head runat="server">
    <title></title>
    <style type="text/css">
        .style1
        {
            font-size: xx-large;
            text-align: center;
            width: 952px;
        }
        .style2
        {
            width: 74%;
        }
        .style3
        {
            font-size: xx-large;
            font-weight: bold;
        }

        .style4
        {
            width: 268435456px;
        }

        .style5
        {
            width: 190px;
        }

        .style6
        {
            width: 163px;
        }
        .style7
        {
            width: 452px;
        }

    </style>
</head>
<body>
    <form id="form1" runat="server">
    <div class="style1">

        <asp:ScriptManager ID="ScriptManager1" runat="server">
        </asp:ScriptManager>
        Early Growth Technology Analysis<br />
    </div>
    <br />
    <table class="style2">
        <tr>
            <td class="style6">
                Seed Term:</td>
            <td class="style7">
                <asp:TextBox ID="txtSeedTerm" runat="server">renewable
energy</asp:TextBox>
            </td>
            <td class="style7">
                &nbsp;</td>
        </tr>
        <tr>
            <td class="style6" >
                Term Generation Sources:</td>
            <td class="style7">
                <asp:CheckBoxList ID="chkboxlistTermGenerationSources" runat="server">
                    <asp:ListItem Selected="True" Value="scirus">Scirus</asp:ListItem>
                    <asp:ListItem Value="compendex">Compendex</asp:ListItem>
                    <asp:ListItem Value="inspec">Inspec</asp:ListItem>
                </asp:CheckBoxList>
            </td>
            <td class="style7">
                <asp:TextBox ID="txtNewTermSource" runat="server">Enter a new
source</asp:TextBox><br />
```

```
                <asp:Button ID="btnAddNewTermSource" runat="server" Text="+"
                    onclick="Button3_Click" />
                <asp:Button ID="btnDeleteTermSource" runat="server" Text="-"
                    onclick="btnDeleteTermSource_Click" />
            </td>
        </tr>
        <tr>
            <td class="style6">
                &nbsp;Term List Size:</td>
            <td class="style7">
                <asp:TextBox ID="txtTermListSize" runat="server">100</asp:TextBox>
            </td>
            <td class="style7">
                &nbsp;</td>
        </tr>
        <tr>
            <td class="style6" >
                Hit Count Source:</td>
            <td class="style7">
                <asp:DropDownList ID="ddlistHitCountSource" runat="server">
                    <asp:ListItem Value="scirushits">Scirus</asp:ListItem>
                    <asp:ListItem Value="compendexhits">Compendex</asp:ListItem>
                    <asp:ListItem Value="inspechits">Inspec</asp:ListItem>
                </asp:DropDownList>
            </td>
            <td class="style7">
                <asp:TextBox ID="txtNewHitSource" runat="server">Enter a new
source</asp:TextBox><br />
                <asp:Button ID="btnAddNewHitSource" runat="server" Text="+"
                    onclick="btnAddNewHitSource_Click" />
                <asp:Button ID="btnDeleteHitSource" runat="server" Text="-"
                    onclick="btnDeleteHitSource_Click" />
            </td>
        </tr>
        <tr>
            <td class="style6" >
                Begin-End Years:</td>
            <td class="style7">
                <asp:TextBox ID="txtBeginYear" runat="server">2006</asp:TextBox>
                <span class="style3">-</span><asp:TextBox ID="txtEndYear"
runat="server">2008</asp:TextBox>
            </td>
            <td class="style7">
                &nbsp;</td>
        </tr>
        <tr>
            <td class="style6">
                Analysis Method:</td>
            <td class="style7">
                &nbsp;&nbsp;&nbsp;&nbsp;&nbsp;&nbsp;&nbsp;
        <asp:DropDownList ID="ddlistAnalysisMethod" runat="server" >
            <asp:ListItem>Total Growth</asp:ListItem>
            <asp:ListItem Selected="True">Early Growth</asp:ListItem>
        </asp:DropDownList>
            </td>
            <td class="style7">
                &nbsp;</td>
        </tr>


        <tr>
            <td class="style6">
                Max # of threads:</td>
            <td class="style7">
                <asp:TextBox ID="txtPOETermNumber" runat="server">10</asp:TextBox>
            </td>
            <td class="style7">
                &nbsp;</td>
        </tr>
        <tr>
            <td class="style6" >
```

```
                Registry:</td>
            <td class="style7">
                <asp:TextBox ID="txtRegistry" runat="server"
Width="307px">http://www.mit.edu/~ayshe/</asp:TextBox>
            </td>
            <td class="style7">
                <asp:TextBox ID="txtMessageBox" runat="server" style="float: left"
                    TextMode="MultiLine" Width="367px"></asp:TextBox>
            </td>
        </tr>
        <tr>
            <td colspan="3" style="text-align: center">

                        <table>
                            <tr>
                                <td style="text-align: center">
                                    <asp:CheckBox ID="chkBoxSave" runat="server"
Text="Save Results" AutoPostBack="True"
                    oncheckedchanged="chkBoxSave_CheckedChanged" />
                                </td>
                                <td style="text-align: center" >
                                    <asp:Button ID="btnGetAll" runat="server"
onclick="Button2_Click"
                    Text="Get All" />
                                </td>
                                <td style="text-align: center" >
                                    &nbsp;</td>
                            </tr>
                            <tr>
                                <td style="text-align: center">
                                    <asp:Button ID="btnAnalyze" runat="server"
onclick="btnAnalyze_Click"
                    style="text-align: center" Text="Get Terms" Width="128px" />
                                    <asp:Button ID="btnRefine" runat="server"
Enabled="False"
                    onclick="btnRefine_Click" Text="Refine" />
                                </td>
                                <td style="text-align: center" >
                                    <asp:Button ID="btnGetHits" runat="server"
onclick="btnHitCounts_Click"
                    Text="Get Hitcounts" Enabled="False" />
                                    <asp:CheckBox ID="chkBoxAddSeedTerm" runat="server"
Text="add seed term" />
                                </td>
                                <td style="text-align: center" >
                                    <asp:Button ID="btnGetTop" runat="server"
onclick="Button1_Click"
                    Text="Get Top Terms" Enabled="False" />
                                </td>
                            </tr>
                            <tr>
                                <td style="text-align: center">
                                    Term List</td>
                                <td style="text-align: center" >
                                    Hit Counts</td>
                                <td style="text-align: center"  >
                                    Top Terms</td>
                            </tr>
                            <tr>
                                <td style="text-align: center">
                                    <asp:ListBox ID="lstBoxTermList" runat="server"
    Height="150px" Width="300px" style="margin-top: 0px"></asp:ListBox>
                                </td>
                                <td style="text-align: center" class="style5">
                                    <asp:ListBox ID="lstBoxHitCounts" runat="server"
Height="150px"
    Width="300px"></asp:ListBox>
                                </td>
                                <td style="text-align: center" class="style4" >
                                    <asp:ListBox ID="lstTopTerms" runat="server"
Height="150px"
```

```
    Width="300px" ></asp:ListBox>
                                </td>
                            </tr>
                            <tr>
                                <td style="text-align: center">
                                    <asp:HyperLink ID="HyperLink1" runat="server"
                    NavigateUrl="~/ResultFiles/termlist.csv"
Visible="False">Download</asp:HyperLink>
                                </td>
                                <td style="text-align: center" >
                                    <asp:HyperLink ID="HyperLink2" runat="server"
                    NavigateUrl="~/ResultFiles/hitcounts.csv"
Visible="False">Download</asp:HyperLink>
                                </td>
                                <td style="text-align: center"  >
                                    <asp:HyperLink ID="HyperLink3" runat="server"
                    NavigateUrl="~/ResultFiles/topterms.csv"
Visible="False">Download</asp:HyperLink>
                                </td>
                            </tr>
                            <tr>
                                <td style="text-align: center">
                                    <asp:Label ID="lblTermsTime" runat="server">Elapsed
time:</asp:Label>
                                </td>
                                <td style="text-align: center">
                                    <asp:Label ID="lblHitCountsTime" runat="server"
Text="Elapsed time:"></asp:Label>
                                </td>
                                <td style="text-align: center" >
                                    <asp:Label ID="lblTopTermsTime" runat="server"
Text="Elapsed time:"></asp:Label>
                                </td>
                            </tr>
                        </table>
                </td>
        </tr>
        </table>

    </form>
</body>
</html>
```

APPENDIX 6 -COMPLETE RANKED TERM LIST FOR SOLAR CASE STUDY

The list below shows the complete results we obtained from our EGTA tool for Solar Energy.

1	Refractometers
2	Nanomagnetics
3	Airfoils
4	Ozonization
5	Reactor Shutdowns
6	Decommissioning (Nuclear Reactors)
7	Polyaniline
8	Nanophotonics
9	Limiters
10	Thermoanalysis
11	Bacteriology
12	Hvdc Power Transmission
13	Photoelectricity
14	Organometallics
15	Graphene
16	Metamaterials
17	Metallizing
18	photovoltaics
19	Vacancies
20	Nanosheets
21	Feedstocks
22	Nuclear Fuel Reprocessing
23	Nanofluidics
24	Holographic Interferometry
25	Aluminum Powder Metallurgy
26	Nanofibers
27	Metallurgy
28	Tribology
29	Feedwater Heaters
30	Photonics
31	Thermoelectric Equipment
32	Exergy
33	Rare Earths
34	Radio Navigation
35	Seismology
36	Heat Exchangers
37	Terahertz Spectroscopy
38	Semiconducting Intermetallics
39	Nanotechnology
40	Bioinformatics
41	Navigation
42	Molds
43	Polyamides
44	Sonochemistry
45	Arctic Engineering
46	Precious Metals
47	Cryogenics
48	Daylighting
49	Microbiology
50	Microelectrodes
51	High Resolution Transmission Electron Microscopy
52	Rockets
53	Microfiltration
54	Composting
55	Escherichia Coli
56	Cytology
57	Mechatronics
58	Robotics
59	Ultrafiltration
60	Fullerenes
61	Quantum Electronics
62	Phosphatases
63	Atomic Force Microscopy
64	Videodisks
65	Squids
66	Nanocomposites
67	Tensors
68	Electromagnets
69	Reflectometers
70	Hose Fittings
71	Feedback
72	Microfluidics
73	Disinfection
74	Biodiesel
75	Oceanography
76	Yeast
77	Planets
78	Maleic Anhydride
79	Thermometers
80	Superconductivity
81	Photoreactivity
82	Biotechnology
83	Photodynamic Therapy
84	Nitric Oxide
85	Cryogenic Equipment
86	Metallorganic Chemical Vapor Deposition
87	Carbon Nanotubes
88	Nanotubes
89	Magnetoelectronics
90	Climatology
91	Phosphorylation
92	Education Computing
93	Electron Optics
94	Coagulation
95	Nitration
96	Meteorology
97	Photometry
98	Dust Collectors
99	Cyclotrons
100	Textiles
101	Ultrasonics
102	Nuclear Medicine
103	Piezoelectricity
104	Biofilms
105	Logistics
106	Radar
107	Global Warming

108	Electron Microscopes
109	Earth (Planet)
110	Pneumatic Drives
111	Pathology
112	Engineering Education
113	Technetium
114	Aluminum Metallurgy
115	Nanoindentation
116	Spice
117	Plasma Diagnostics
118	Wind Turbines
119	Fossil Fuels
120	Oxidative Stress
121	Ozone Water Treatment
122	Polyesters
123	Genes
124	Spacecraft
125	Nanostructured Materials
126	Bioremediation
127	Metadata
128	Musical Instruments
129	Metabolism
130	Batch Reactors
131	Tin
132	Solar Water Heaters
133	Photorefractive Crystals
134	Atmospheric Aerosols
135	Plastics
136	Tellurium Compounds
137	Uranium
138	Biogeochemistry
139	Biochemistry
140	Xylose
141	Heterojunction Bipolar Transistors
142	Dryers (Equipment)
143	Optical Microscopy
144	Wine
145	Solar Cells
146	Bioreactors
147	Transmission Electron Microscopy
148	Dosimetry
149	Oligonucleotides
150	Superfluid Helium
151	Polyelectrolytes
152	Ozone
153	Transuranium Elements
154	Robots
155	Ion Microscopes
156	Nanowires
157	Peptides
158	Greenhouse Gases
159	Acoustics
160	Compressors
161	Spectrum Analyzers
162	Bioactivity
163	Microelectronics
164	Fiber Bragg Gratings
165	Electric Switchgear
166	Energy Policy
167	Pneumatics
168	Antibiotics
169	Sales
170	Renewable Energy Resources
171	Electromagnetism
172	Nanocrystals
173	Surveying
174	Photocatalysis
175	Nuclear Fuels
176	Forestry
177	Nuclear Industry
178	Electrocatalysis
179	Nitriding
180	Bioconversion
181	Gold Metallurgy
182	Satellites
183	Nickel Alloys
184	Reentry
185	Monolayers
186	Quantum Chemistry
187	Mems
188	Automation
189	Tropics
190	Stars
191	Galaxies
192	Gadolinium
193	Rna
194	Oxygenation
195	Pneumatic Control Equipment
196	Dna
197	Solar Power Generation
198	Magnetism
199	Quantum Optics
200	Photoelectrochemical Cells
201	Coal Industry
202	Computer Crime
203	Abstracting
204	Interoperability
205	Nuclear Reactors
206	Silicon Solar Cells
207	Solar Energy
208	Wimax
209	Hydroxylation
210	Siphons
211	Scintillation
212	Solar Concentrators
213	Storms
214	Gas Emissions
215	Corrosion Protection
216	Microwaves
217	Radio Astronomy
218	Silver
219	Electrochemistry
220	Mirrors
221	Rotors
222	Sols
223	Multiobjective Optimization
224	Supply Chain Management
225	Strontium
226	Gas Turbines
227	Virtual Reality
228	Selenium
229	Nuclear Engineering
230	Unsaturated Polymers
231	Economics

232	Smart Antennas
233	Ethanol Fuels
234	Radioactive Waste Disposal
235	Ecosystems
236	Solar Power Plants
237	Wetlands
238	Cellulosic Ethanol
239	Purification
240	Engines
241	Mass Spectrometry
242	Feedback Control
243	Scanning Electron Microscopy
244	Arsenic
245	Plants (Botany)
246	State Feedback
247	Biology
248	Crack Tips
249	Balloons
250	High Performance Liquid Chromatography
251	Climate Change
252	Fuels
253	Bacteria
254	Pesticides
255	Cmos Integrated Circuits
256	Enzymes
257	Plutonium
258	Radioactivity
259	Wireless Sensor Networks
260	Sun
261	Teaching
262	Interferometry
263	Coal
264	Nanorods
265	Image Coding
266	Chitin
267	Flywheels
268	Human Engineering
269	Safety Engineering
270	Catalysts
271	Information Science
272	Probes
273	Underwater Acoustics
274	Solar System
275	Glucose
276	Polymers
277	Seawater
278	Spectrometry
279	Photonic Crystals
280	Biodegradation
281	Agriculture
282	Boiler Control
283	Transparency
284	Neurons
285	Aerodynamics
286	Recycling
287	Gels
288	Nuclear Energy
289	Platinum
290	Amino Acids
291	Superconducting Films
292	Tunneling (Excavation)
293	Welding
294	Cell Culture
295	Wind Tunnels
296	Rocket Engines
297	Barium
298	Transportation
299	Turbines
300	Mooring
301	Packaging Materials
302	Ferroelectric Films
303	Reheat Cycle
304	Historic Preservation
305	Microlenses
306	Styrene
307	Corrosion Resistance
308	Gold
309	Chromatography
310	Audio Systems
311	Turbomachine Blades
312	Solar Collectors
313	Counting Circuits
314	Biosensors
315	Plating
316	Intermetallics
317	Air Pollution
318	Solar Equipment
319	Scintillation Counters
320	Authentication
321	Caissons
322	Steam
323	Internet
324	Lakes
325	Pumps
326	Block Copolymers
327	Rapid Thermal Annealing
328	Cell Membranes
329	Fluorescence
330	Sustainable Development
331	Energy Resources
332	Harvesting
333	Glycerol
334	Enzyme Activity
335	Technology
336	Radio
337	Modulators
338	Urea
339	Optical Resonators
340	Nucleic Acids
341	Biomass
342	Metabolites
343	Digital Cameras
344	Commerce
345	Diamonds
346	Remote Sensing
347	Printing
348	Artificial Intelligence
349	Greenhouses
350	Nuclear Physics
351	Flow Visualization
352	Roofs
353	Proteins
354	Hydrogen Peroxide
355	Research
356	Modems

357	Photocatalysts
358	Fermentation
359	Ice
360	Corrosion
361	Single-Walled Carbon Nanotubes (Swcn)
362	Nanoparticles
363	Automobiles
364	Polypyrroles
365	Hydrophobicity
366	Gravitation
367	Membrane Technology
368	Nanostructures
369	Calcium
370	Electroosmosis
371	Hydrogels
372	Hydraulics
373	Ion Exchangers
374	Fish
375	Crosstalk
376	Particle Detectors
377	Semiconductor Quantum Dots
378	Management
379	Towers
380	Ultrathin Films
381	Coatings
382	Phospholipids
383	Rivers
384	Membranes
385	Water Resources
386	Pressure Vessels
387	Mesfet Devices
388	Holmium
389	Surface Plasmon Resonance
390	Biogas
391	Curricula
392	Hydroxyapatite
393	Ultrashort Pulses
394	Superlattices
395	Nematic Liquid Crystals
396	Neodymium Lasers
397	Ytterbium
398	Computer Science
399	Solar Refrigeration
400	Maltose
401	Pollution
402	Fatty Acids
403	Landing
404	Transesterification
405	Porphyrins
406	Energy Management
407	Ferritic Steel
408	Gate Dielectrics
409	Hydraulic Machinery
410	Imaging Systems
411	Optical Multilayers
412	Resonant Tunneling Diodes
413	Superconducting Materials
414	Thermodynamics
415	Composite Micromechanics
416	Copolymers
417	Alloys
418	Solid Oxide Fuel Cells (Sofc)
419	Wind Power
420	Microorganisms
421	Clarification
422	Lasers
423	Adhesion
424	Zinc
425	Potentiometers (Electric Measuring Instruments)
426	Sapphire
427	Antigen-Antibody Reactions
428	Nuclear Power Plants
429	Liquid Chromatography
430	Fuzzy Neural Networks
431	Farms
432	Nonlinear Optics
433	Software Engineering
434	Cryptography
435	Gas Chromatography
436	Accidents
437	Pixels
438	Turbidity
439	Petri Nets
440	Quality Assurance
441	Mortar
442	Electric Industry
443	Fiber Optics
444	Plastic Products
445	Lithium Batteries
446	Steam Engineering
447	Self Assembled Monolayers
448	Denitrification
449	Isotopes
450	Cadmium Compounds
451	Microbial Fuel Cells
452	Fuel Cells
453	Superconducting Wire
454	Conservation
455	Mesoporous Materials
456	Acids
457	Recovery
458	Galvanomagnetic Effects
459	Electricity
460	Buildings
461	Sol-Gels
462	Mechanical Engineering
463	Biofilters
464	Multilayers
465	Color
466	Stainless Steel
467	Security Systems
468	Computer Aided Design
469	Cells
470	Controllers
471	Network Security
472	Water Content
473	Enclosures
474	Drug Delivery
475	Ferroelectricity
476	Pipe
477	Resonators
478	Presses (Machine Tools)
479	Magnetohydrodynamics
480	Forecasting

481	High Temperature Superconductors
482	Detectors
483	Defects
484	Bicmos Technology
485	Ionizing Radiation
486	Biochemical Engineering
487	Military Applications
488	Materials Science
489	Photoelectron Spectroscopy
490	Boilers
491	Learning Systems
492	Fluid Dynamics
493	Sedimentology
494	Power Electronics
495	Microchannels
496	Fracture
497	Spectrometers
498	Sensor Networks
499	Repair
500	Wireless Networks
501	Copper
502	Fault Tolerance
503	Mesh Networking
504	Water Recycling
505	Plasmons
506	Particle Swarm Optimization (Pso)
507	Scanning
508	Iridium
509	Paramagnetic Resonance
510	World Wide Web
511	Grid Computing
512	Sodium
513	Bridges
514	Standards
515	Data Mining
516	Image Analysis
517	Rheology
518	Planning
519	Photovoltaic Cells
520	Landforms
521	Computer Architecture
522	Soil Pollution
523	Vlsi Circuits
524	Linearization
525	Plasmas
526	Mesh Generation
527	Laminar Flow
528	Raman Spectroscopy
529	Ceramic Materials
530	Hydrodynamics
531	Magnets
532	Air Conditioning
533	Hybrid Materials
534	Polychlorinated Biphenyls
535	Nuclear Magnetic Resonance Spectroscopy
536	Blending
537	Crops
538	Yttria Stabilized Zirconia
539	Resonance
540	Design
541	Phosphorus
542	Soil Mechanics
543	Krypton
544	Catalysis
545	Ethanol
546	Synthetic Fuels
547	Steam Power Plants
548	Wastewater
549	Polystyrenes
550	Neodymium
551	Capillary Flow
552	Sugar Substitutes
553	Voltage Regulators
554	Luminescence
555	Doppler Effect
556	Water Pollution
557	Scheduling
558	Ternary Systems
559	Strategic Planning
560	Retail Stores
561	Project Management
562	Water Heaters
563	Superparamagnetism
564	Chromatographic Analysis
565	Cultivation
566	Microelectromechanical Devices
567	Fast Reactors
568	Wells
569	Glass
570	Hazards
571	Competition
572	Electrocatalysts
573	Table Lookup
574	Profitability
575	Sodium Sulfate
576	Heterojunctions
577	Hardware
578	Customer Satisfaction
579	Fuel Economy
580	Distillation
581	Matlab
582	Security Of Data
583	Cables
584	Pyrolysis
585	Computational Fluid Dynamics
586	Software Design
587	Photoluminescence Spectroscopy
588	Transmitters
589	Optical Waveguides
590	Extraction
591	Wood Products
592	Vegetable Oils
593	School Buildings
594	Scanning Tunneling Microscopy
595	Risk Management
596	Photosensitizers
597	Product Design
598	Waveguides
599	Silicate Minerals
600	Gas Fuel Purification
601	Waste Management
602	Radiation
603	Electric Utilities

604	Hydrographic Surveys
605	Iron
606	Electrolytes
607	Quality Of Service
608	Electric Vehicles
609	Water Management
610	Laws And Legislation
611	Neural Networks
612	Lithium
613	Fuzzy Logic
614	Targets
615	Free Energy
616	Thermoplastics
617	Molecular Orientation
618	Testing
619	Electrodeposition
620	Laser Ablation
621	Biological Materials
622	Photodegradation
623	Soils
624	Accelerometers
625	Multiphoton Processes
626	Delta Sigma Modulation
627	Gas Plants
628	Carbon Steel
629	Materials
630	Anaerobic Digestion
631	Ducts
632	Gamma Ray Spectrometers
633	Power Plants
634	Computer Software
635	Lanthanum
636	Manganese
637	Metals
638	Apatite
639	Potassium
640	Semiconductor Lasers
641	Wire
642	Molecular Dynamics
643	Reynolds Number
644	Drug Products Plants
645	Methane
646	Chemical Sensors
647	Communication
648	Free Radicals
649	Paint
650	Molybdenum
651	Creep
652	Fabrics
653	Knowledge Management
654	Production Engineering
655	Biological Membranes
656	Hydration
657	Nutrients
658	Solar Heating
659	Visual Communication
660	Network Architecture
661	Lipid Bilayers
662	Dust
663	Dissociation
664	Natural Gas
665	Military Equipment
666	Cast Iron
667	Multiprocessing Systems
668	Translation (Languages)
669	Spectroscopy
670	Lenses
671	Silicon
672	Air
673	Gas Detectors
674	Quartz Crystal Microbalances
675	Rain
676	Dyes
677	Solar Buildings
678	Morphology
679	Fuzzy Systems
680	Tritium
681	Water
682	Water Conservation
683	Catchments
684	Uranium Dioxide
685	Maintenance
686	Rutherford Backscattering Spectroscopy
687	Thermonuclear Reactions
688	Josephson Junction Devices
689	Public Utilities
690	Induction Motors
691	Gas Hydrates
692	Gasoline
693	Energy Conservation
694	Gas Generators
695	Backscattering
696	Irrigation
697	Wheels
698	Machinery
699	Field Emission Microscopes
700	Refractive Index
701	Emulsification
702	Transceivers
703	Terbium Alloys
704	Nuclear Propulsion
705	Solar Radiation
706	Software Testing
707	Fracture Mechanics
708	Hydrocarbons
709	Cobalt
710	Waste Incineration
711	Oxide Minerals
712	Radiation Protection
713	Fibers
714	Sieves
715	Gases
716	Industrial Applications
717	Combustion
718	Climate Models
719	Pulsed Laser Deposition
720	Buses
721	X Ray Photoelectron Spectroscopy
722	Toxicity
723	Photolithography
724	Investments
725	Diesel Engines
726	Computer Programming
727	Biodegradable Polymers

728	Semiconductor Materials
729	Organic Light Emitting Diodes (Oled)
730	Wastewater Treatment
731	Copolymerization
732	Backpropagation
733	Eutrophication
734	Interactive Computer Systems
735	Electric Energy Storage
736	Microwave Integrated Circuits
737	Offshore Structures
738	Modulation
739	Electrolytic Capacitors
740	Masks
741	Image Segmentation
742	Genetic Algorithms
743	Europium
744	Organic Chemicals
745	Automobile Engines
746	Shrinkage
747	Ultraviolet Radiation
748	Floors
749	Feeding
750	Photoionization
751	Program Compilers
752	Tidal Power
753	Chitosan
754	Curing
755	Nonlinear Feedback
756	Industrial Plants
757	Carbon Dioxide
758	Remediation
759	Developing Countries
760	Spectroscopic Analysis
761	Phosphorescence
762	Ruthenium
763	Biocompatibility
764	Biochemical Fuel Cells
765	Luminous Paint
766	Monitoring
767	Flexible Manufacturing Systems
768	Photoluminescence
769	Poles
770	Hydrolysis
771	Energy Storage
772	Public Works
773	Embedded Software
774	Water Filtration
775	Computer Hardware
776	Fossil Fuel Power Plants
777	Transient Analysis
778	Fuel Storage
779	Xenon
780	Ac Motors
781	Fluidized Beds
782	Effluents
783	Fluids
784	Plastic Coatings
785	Electric Rectifiers
786	Fracture Toughness
787	Multicasting
788	Water Quality
789	Bayesian Networks
790	Aluminum Alloys
791	Flocculation
792	Hybrid Computers
793	Inspection
794	Acetylene
795	Throughput
796	Aluminum
797	Game Theory
798	Ultra-Wideband (Uwb)
799	Electric Machinery
800	Evapotranspiration
801	Waves
802	Photovoltaic Effects
803	Verification
804	Perovskite
805	Computer Simulation
806	Ductility
807	Evolutionary Algorithms
808	Cellular Neural Networks
809	Coolants
810	Nitrification
811	Air Quality
812	Energy Efficiency
813	Pressure Transducers
814	Gallium
815	Annealing
816	Optoelectronic Devices
817	Moisture
818	Quantum Well Lasers
819	Reinforced Concrete
820	Electrochemical Corrosion
821	Office Buildings
822	Magnetic Susceptibility
823	Adatoms
824	Photopolymerization
825	Ocean Currents
826	Nucleate Boiling
827	Fluidization
828	Tall Buildings
829	Deuterium
830	Gold Deposits
831	Physical Optics
832	Vehicles
833	Soil Conservation
834	Solvent Extraction
835	Irradiation
836	Sodium Chloride
837	X Rays
838	Sewage
839	Learning Algorithms
840	Sulfur
841	Stress Analysis
842	Chromium
843	Light Emission
844	Power Converters
845	Cements
846	Photonic Band Gap
847	Chirality
848	Speed
849	Soil Moisture
850	Optical Instruments
851	Protective Coatings

852	Intelligent Buildings
853	Molecules
854	Stiffness
855	Automata Theory
856	Olefins
857	Boron
858	Kinematics
859	Semiconductor Growth
860	Polarons
861	Costs
862	Lactic Acid
863	Fused Silica
864	Fusion Reactors
865	Graphite
866	Polyethylenes
867	Reliability
868	Portable Equipment
869	Esterification
870	Polysaccharides
871	Lanthanum Alloys
872	Resins
873	Nitrogen Plasma
874	Risk Assessment
875	Radioisotopes
876	Hardness
877	Atomic Spectroscopy
878	Excitons
879	Ion Exchange Membranes
880	Retrofitting
881	Specifications
882	Standardization
883	Zirconia
884	Copper Compounds
885	Amorphous Alloys
886	Environmental Impact
887	Chemisorption
888	Dispersions
889	Molecular Beam Epitaxy
890	Light
891	Cellulose
892	Cosmic Ray Detectors
893	Gallium Alloys
894	Internet Telephony
895	Fabrication
896	Reaction Kinetics
897	Telecommunication Networks
898	Leaching
899	Synthesis (Chemical)
900	Networks (Circuits)
901	Magnesium
902	Redox Reactions
903	Land Use
904	Infrared Detectors
905	Sensors
906	Supply Chains
907	Antiferromagnetism
908	Critical Current Density (Superconductivity)
909	Gallium Arsenide
910	Silicon Carbide
911	Thermodynamic Properties
912	Liquids
913	Microstructure
914	Electric Clocks
915	Fuzzy Control
916	Groundwater
917	Dynamics
918	Oxygen
919	Ph Sensors
920	Photodetectors
921	Adsorption
922	Schottky Barrier Diodes
923	Magnetic Materials
924	Rhodium
925	Crosslinking
926	Methanol Fuels
927	Millimeter Wave Devices
928	Electrochemical Impedance Spectroscopy
929	Semiconducting Selenium Compounds
930	Automobile Parts And Equipment
931	Fiber Optic Sensors
932	Aquifers
933	Roads And Streets
934	Microwave Devices
935	High Temperature Reactors
936	Plasma Jets
937	Tensile Strength
938	Tuning
939	Runoff
940	Machine Design
941	Models
942	Micelles
943	Optimization
944	Water Treatment
945	Microwave Measurement
946	Semiconducting Gallium
947	Photolysis
948	Ultraviolet Photoelectron Spectroscopy
949	Solar Absorbers
950	Semiconductor Quantum Wells
951	Reinforcement
952	Infrared Spectroscopy
953	Carbon Fibers
954	Radioactive Wastes
955	Quantum Theory
956	Routers
957	Laser Beams
958	Capillarity
959	Spray Pyrolysis
960	Formic Acid
961	Electron Guns
962	Cesium
963	Molecular Oxygen
964	Ultraviolet Spectroscopy
965	Resonant Tunneling
966	Helium
967	Electroluminescence
968	Antennas
969	Strontium Alloys
970	Wireless Telecommunication Systems
971	Speech Recognition
972	Powders
973	Iron Alloys

974	Heat Transfer
975	Broadband Amplifiers
976	Electric Power Generation
977	Gamma Rays
978	Welds
979	Stabilization
980	Piezoelectric Actuators
981	Process Engineering
982	Contamination
983	Environmental Regulations
984	Quality Control
985	Electric Power Transmission
986	Water Waves
987	Magnetron Sputtering
988	Methanol
989	Thermoelectric Power Plants
990	Crystals
991	Phosphorus Compounds
992	Microporosity
993	Reservoirs (Water)
994	Paraffins
995	Ethylene
996	Telecommunication Systems
997	Diesel Fuels
998	Negative Resistance
999	Pulsed Lasers
1000	Induction Heating
1001	Cracks
1002	Acetone
1003	Heat Resistance
1004	Neutrons
1005	Quantum Well Infrared Photodetectors
1006	Film Growth
1007	Optical Materials
1008	Epitaxial Films
1009	Humidity Sensors
1010	Degradation
1011	Gallium Nitride
1012	Palladium
1013	Electric Generators
1014	Platinum Alloys
1015	Circuit Simulation
1016	Photocurrents
1017	Durability
1018	Lead
1019	Electrolytic Cells
1020	Orbital Laboratories
1021	Mobile Ad Hoc Networks
1022	Microwave Irradiation
1023	District Heating
1024	Ph Effects
1025	Winding
1026	Optical Fibers
1027	Hybrid Systems
1028	Fading Channels
1029	Chromophores
1030	Mechanisms
1031	Electrochemical Cells
1032	Topology
1033	Substrates
1034	Electrospray Ionization
1035	Sheet Metal
1036	Access Control
1037	Water Supply
1038	Radiation Effects
1039	Photoresists
1040	Waste Heat
1041	Global Optimization
1042	Crystallization
1043	Silica
1044	Geologic Models
1045	Carbon Films
1046	Errors
1047	Nitrogen
1048	Electric Power Systems
1049	Time Series
1050	Thiophene
1051	Diffusion
1052	Feature Extraction
1053	Piezoelectric Materials
1054	Deposition
1055	Mobile Telecommunication Systems
1056	Polycondensation
1057	Flue Gases
1058	Electric Batteries
1059	Solar Cell Arrays
1060	Potable Water
1061	Complexation
1062	Laser Theory
1063	Organic Compounds
1064	Feedback Amplifiers
1065	Mathematical Programming
1066	Experiments
1067	Underwater Soils
1068	Water Tanks
1069	Binding Energy
1070	Composite Membranes
1071	X Ray Diffraction
1072	Wind Effects
1073	Oxidation
1074	Digital Control Systems
1075	Plasma Applications
1076	Glass Lasers
1077	Systems Analysis
1078	Sewage Treatment
1079	Polymer Blends
1080	Augmented Reality
1081	Sugar (Sucrose)
1082	Thermogravimetric Analysis
1083	Radio Communication
1084	Superconducting Magnets
1085	Laser Pulses
1086	Polymer Films
1087	Cooling Systems
1088	Water Cooling Systems
1089	Spontaneous Emission
1090	Energy Gap
1091	Phase Modulation
1092	Silicon Alloys
1093	Cascades (Fluid Mechanics)
1094	Ionic Liquids
1095	Pressure Sensors
1096	Thermocouples

1097	Ito Glass
1098	Lanthanum Compounds
1099	Aldehydes
1100	Magnetic Fields
1101	Fourier Transform Infrared Spectroscopy
1102	Magnetic Films
1103	Decision Making
1104	Antimony
1105	Dc-Dc Converters
1106	Domestic Appliances
1107	Stoichiometry
1108	Catalytic Oxidation
1109	Actuators
1110	Environmental Protection Agency
1111	Personal Computers
1112	Embedded Systems
1113	Ac Generator Motors
1114	Crystal Growth
1115	Integrated Circuit Testing
1116	Direct Methanol Fuel Cells (Dmfc)
1117	Compressed Air
1118	Water Treatment Plants
1119	Wafer Bonding
1120	Niobium
1121	Silicides
1122	Electric Grounding
1123	Image Processing
1124	Wind Stress
1125	Human Computer Interaction
1126	Millimeter Waves
1127	Coastal Zones
1128	Materials Handling
1129	Pumping Plants
1130	Plasma Stability
1131	Optical Films
1132	Hydrogen
1133	Aluminum Coatings
1134	Reaction Intermediates
1135	Power Generation
1136	Electric Converters
1137	Quenching
1138	Blooms (Metal)
1139	Atmospheric Humidity
1140	Surface Chemistry
1141	X Ray Spectroscopy
1142	Solidification
1143	Bearings (Structural)
1144	Ferromagnetism
1145	Cyclic Voltammetry
1146	Lithium Compounds
1147	Navier Stokes Equations
1148	Radio Waves
1149	Heavy Metals
1150	Spin Glass
1151	Contact Angle
1152	Buoyancy
1153	Plasma Deposition
1154	Scattering
1155	Oxygen Vacancies
1156	Organic Polymers
1157	Thin Films
1158	Calibration
1159	Torque
1160	Activation Energy
1161	Grafting (Chemical)
1162	Printed Circuit Boards
1163	Stress Concentration
1164	Telecommunication Equipment
1165	Monolithic Microwave Integrated Circuits
1166	Phosphors
1167	Reluctance Motors
1168	Liquid Crystals
1169	Stresses
1170	Water Vapor
1171	Programmable Logic Controllers
1172	Heating Equipment
1173	Film Preparation
1174	Digital Devices
1175	Semiconducting Organic Compounds
1176	Heat Radiation
1177	Semiconducting Aluminum Compounds
1178	Underground Cables
1179	Energy Utilization
1180	Dissolution
1181	Gas Adsorption
1182	Wavelet Transforms
1183	Mercury (Metal)
1184	Body Fluids
1185	Sensitivity Analysis
1186	Cadmium Alloys
1187	Active Filters
1188	Reliability Analysis
1189	Oxidation Resistance
1190	Pilot Plants
1191	Plasma Display Devices
1192	Ferrite
1193	Feedforward Neural Networks
1194	Linear Programming
1195	Organic Pollutants
1196	Geometry
1197	Epitaxial Growth
1198	Fault Tolerant Computer Systems
1199	Stochastic Systems
1200	Ray Tracing
1201	Measurements
1202	Sulfur Dioxide
1203	Cerium
1204	Isomers
1205	Polarization
1206	Pore Size
1207	Domain Walls
1208	Clustering Algorithms
1209	Gold Coatings
1210	Calorific Value
1211	Partial Differential Equations
1212	Classifiers
1213	Titanium
1214	Desorption
1215	Electron Tubes
1216	Dynamical Systems

1217	Bandwidth
1218	Industrial Water Treatment
1219	Vacuum
1220	Aromatic Hydrocarbons
1221	Ad Hoc Networks
1222	High Energy Physics
1223	Heat Pump Systems
1224	Superconducting Cables
1225	Physisorption
1226	Grain Growth
1227	Reduction
1228	Point Defects
1229	Tunnel Junctions
1230	Frequency Synthesizers
1231	Topography
1232	Sound Reproduction
1233	Budget Control
1234	Secondary Ion Mass Spectrometry
1235	Electrochemical Sensors
1236	Jitter
1237	Computer Operating Systems
1238	Organic Acids
1239	Electric Wire
1240	Bandpass Filters
1241	Lithium Alloys
1242	Light Transmission
1243	Transients
1244	Silicates
1245	Manganese Oxide
1246	Field Emission
1247	Photoexcitation
1248	Philosophical Aspects
1249	Pulse Modulation
1250	Stress Corrosion Cracking
1251	Optical Fiber Fabrication
1252	Spheres
1253	Crystal Structure
1254	Nucleation
1255	Coercive Force
1256	Signal Detection
1257	Electric Inverters
1258	Heavy Water
1259	Structural Design
1260	Organic Carbon
1261	Materials Properties
1262	Data Transfer
1263	Vapors
1264	Zinc Oxide
1265	Mass Transfer
1266	Heating
1267	Temperature
1268	Logic Design
1269	Energy Transfer
1270	Glassy Carbon
1271	Galerkin Methods
1272	Computational Methods
1273	Hydrogen Production
1274	Semiconducting Silicon
1275	Surface Morphology
1276	Electrochemical Properties
1277	Ionization
1278	Electric Network Analysis
1279	Ion Exchange
1280	Molecular Beams
1281	Groundwater Pollution
1282	Algorithms
1283	Image Sensors
1284	Sun Hoods
1285	Metal Recovery
1286	Salts
1287	Pulsed Laser Applications
1288	Total Quality Management
1289	Stability
1290	Removal
1291	Permanent Magnets
1292	Quantum Efficiency
1293	Electronic Guidance Systems
1294	Support Vector Machines
1295	Oscillators (Electronic)
1296	Vacuum Technology
1297	Reflection
1298	Sensor Nodes
1299	Oxygen Sensors
1300	Estimation
1301	Analog Circuits
1302	Friction
1303	Solids
1304	Self Assembly
1305	Atmospheric Temperature
1306	Emission Spectroscopy
1307	Stators
1308	Ordinary Differential Equations
1309	Digital Integrated Circuits
1310	Hall Mobility
1311	Surface Waters
1312	Dissolved Oxygen Sensors
1313	Propylene
1314	Fuzzy Sets
1315	Salinity Measurement
1316	Pressure Effects
1317	Humidity Control
1318	Optical Systems
1319	Optical Kerr Effect
1320	Protonation
1321	Electric Inductors
1322	Power Amplifiers
1323	Tungsten
1324	Waste Disposal
1325	Failure Analysis
1326	Data Processing
1327	Cerium Alloys
1328	Grain (Agricultural Product)
1329	Ketones
1330	Ion Exchange Resins
1331	Predictive Control Systems
1332	Differential Equations
1333	Functional Groups
1334	Dewatering
1335	Cooling
1336	Emission Control
1337	Semiconductor Devices
1338	Atmospheric Pressure
1339	Stochastic Control Systems
1340	Titanium Alloys

1341	Absorption
1342	Control System Analysis
1343	Specific Heat
1344	Temperature Control
1345	Metal Analysis
1346	Oxygen Supply
1347	Conformal Mapping
1348	Energy Conversion
1349	Photodiodes
1350	Brittle Fracture
1351	Alkalinity
1352	Chemical Water Treatment
1353	Temperature Sensors
1354	Biological Water Treatment
1355	Switching
1356	Combinatorial Optimization
1357	Esters
1358	Industrial Waste Treatment
1359	Optical Sensors
1360	Scheduling Algorithms
1361	Monomers
1362	Glow Discharges
1363	Ventilation Exhausts
1364	Torque Control
1365	Metal Working
1366	Differential Scanning Calorimetry
1367	Model Predictive Control
1368	Nonlinear Systems
1369	Doping (Additives)
1370	Concurrency Control
1371	Mach Number
1372	Anodes
1373	Electrodes
1374	Vectors
1375	Regression Analysis
1376	Quartz
1377	Polymerization
1378	Potential Energy
1379	Moisture Control
1380	Light Water Reactors
1381	Agglomeration
1382	Elementary Particles
1383	Sintering
1384	Thermal Conductivity Of Gases
1385	Hafnium
1386	Germanium
1387	Acoustic Waves
1388	Manganese Compounds
1389	Economic And Social Effects
1390	Laser Excitation
1391	Model Buildings
1392	Mechanical Properties
1393	Optical Properties
1394	Heat Sinks
1395	Porous Silicon
1396	Turbulence Models
1397	Stearic Acid
1398	Fluorine
1399	Semiconductor Junctions
1400	Green'S Function
1401	Inductance
1402	Flow Of Fluids
1403	Optical Constants
1404	Problem Solving
1405	Indium Sulfide
1406	Message Passing
1407	Plastic Films
1408	Valves (Mechanical)
1409	Composite Films
1410	Oxide Films
1411	Adaptive Control Systems
1412	Zinc Sulfide
1413	Capacitors
1414	Silicon Wafers
1415	Integrated Circuits
1416	Indium
1417	Thermal Conductivity
1418	Rapid Thermal Processing
1419	Organic Conductors
1420	Chemical Oxygen Demand
1421	Excited States
1422	Computer Control Systems
1423	Condensation
1424	Backpropagation Algorithms
1425	Reactive Power
1426	Data Communication Systems
1427	Dc Power Transmission
1428	Network Protocols
1429	Decision Theory
1430	Bipolar Transistors
1431	Organic Coatings
1432	Semiconducting Indium
1433	Cost Effectiveness
1434	Interconnection Networks
1435	Crossbar Equipment
1436	Silver Alloys
1437	Real Time Systems
1438	Data Flow Analysis
1439	Anisotropy
1440	Coalescence
1441	Cobalt Compounds
1442	Liquid Nitrogen
1443	Electrons
1444	Wave Power
1445	Digital Image Storage
1446	User Interfaces
1447	Metastable Phases
1448	Electronic Structure
1449	Leakage Currents
1450	Electromagnetic Waves
1451	Cold Storage
1452	Magnetic Properties
1453	Iron Compounds
1454	Semiconductor Diodes
1455	Zirconium
1456	Selective Catalytic Reduction
1457	Beam Plasma Interactions
1458	Organic Solvents
1459	Spin Dynamics
1460	Natural Convection
1461	Semiconducting Silicon Compounds
1462	Error Correction

1463	Heat Convection
1464	Ground State
1465	Heat Storage
1466	Ionization Chambers
1467	Dynamic Programming
1468	Mathematical Models
1469	Ion Beams
1470	Water Supply Systems
1471	Iridium Compounds
1472	Engine Cylinders
1473	Light Polarization
1474	Molybdenum Oxide
1475	Single Crystals
1476	Data Structures
1477	Electronic Properties
1478	Frequency Modulation
1479	Electron Gas
1480	Process Control
1481	Surface Reactions
1482	Semiconducting Lead Compounds
1483	Routing Algorithms
1484	Boron Nitride
1485	Multilayer Films
1486	Fuzzy Inference
1487	Liquid Films
1488	Diodes
1489	Cantilever Beams
1490	Pulse Amplitude Modulation
1491	Numerical Analysis
1492	Thermodynamic Stability
1493	Spectrum Analysis
1494	Capacitance
1495	Decoding
1496	Robust Control
1497	Amorphous Films
1498	Electric Drives
1499	Temperature Measurement
1500	Non Newtonian Flow
1501	Activated Carbon
1502	Signal Processing
1503	Cathodes
1504	Formal Methods
1505	Time Series Analysis
1506	Nuclear Reactor Accidents
1507	Charge Transfer
1508	Nusselt Number
1509	Electromagnetic Fields
1510	Light Emitting Diodes
1511	Catalyst Selectivity
1512	Design Of Experiments
1513	Flash Memory
1514	Proton Exchange Membrane Fuel Cells (Pemfc)
1515	Flow Control
1516	Time Measurement
1517	Encoding (Symbols)
1518	Ethylene Glycol
1519	Voltage Control
1520	Solenoids
1521	Time Domain Analysis
1522	Transport Properties
1523	Polynomials
1524	Electron Mobility
1525	Numerical Methods
1526	Suspensions (Fluids)
1527	Adsorption Isotherms
1528	Resource Allocation
1529	Velocity
1530	Tubes (Components)
1531	Phonons
1532	Normal Distribution
1533	Transistors
1534	Phase Change Materials
1535	Alcohols
1536	Electric Potential
1537	Cracking (Chemical)
1538	Thermal Energy
1539	Flow Of Gases
1540	Dielectric Properties
1541	Pressure Drop
1542	Coding Errors
1543	Crack Propagation
1544	Amorphous Silicon
1545	Internet Protocols
1546	Nitrides
1547	Electron Beams
1548	Harmonic Analysis
1549	Forced Convection
1550	Light Sources
1551	Dielectric Materials
1552	Phenols
1553	Sulfur Compounds
1554	Absorption Spectroscopy
1555	Glass Transition
1556	Plates (Structural Components)
1557	Adaptive Algorithms
1558	Ions
1559	Protons
1560	Electrochemical Electrodes
1561	Activated Sludge Process
1562	Structural Properties
1563	Surface Testing
1564	Solubility
1565	Water Absorption
1566	Surfaces
1567	Leakage (Fluid)
1568	Electric Fields
1569	Amplitude Modulation
1570	Zinc Compounds
1571	Tantalum
1572	Photons
1573	Functional Polymers
1574	Polymeric Films
1575	Ion Implantation
1576	Conducting Polymers
1577	Passive Networks
1578	Thermal Expansion
1579	Sol-Gel Process
1580	Catalyst Supports
1581	System Stability
1582	Pumping (Laser)
1583	Timing Jitter
1584	Metallic Compounds
1585	Traffic Surveys
1586	Passive Solar Buildings

1587	Order Disorder Transitions
1588	Packet Loss
1589	Program Processors
1590	Heat Flux
1591	Multi Agent Systems
1592	Three Dimensional
1593	Dissolved Oxygen
1594	Proton Conductivity
1595	Telecommunication Traffic
1596	Catalyst Poisoning
1597	Atom Transfer Radical Polymerization
1598	Frequency Converters
1599	Size Distribution
1600	Argon
1601	Electric Resistance
1602	Sulfuric Acid
1603	Variational Techniques
1604	Velocity Measurement
1605	Model Checking
1606	Rotation
1607	Catalyst Activity
1608	Markov Processes
1609	Gas Absorption
1610	Operating Costs
1611	Hafnium Compounds
1612	Free Radical Reactions
1613	Digital Signal Processing
1614	Water Distribution Systems
1615	Sputtering
1616	Structural Optimization
1617	Strontium Compounds
1618	Carrier Mobility
1619	Satellite Communication Systems
1620	Laser Heating
1621	Amorphous Carbon
1622	Audio Frequency Amplifiers
1623	Finite Element Method
1624	Oxides
1625	Polymeric Membranes
1626	Phosphoric Acid
1627	Nitrogen Compounds
1628	Barium Compounds
1629	Electric Discharges
1630	Probability
1631	Reforming Reactions
1632	Yttrium Barium Copper Oxides
1633	Ammonium Compounds
1634	Mathematical Operators
1635	Mixed Convection
1636	Porous Materials
1637	Titanium Dioxide
1638	Paper Sheeting
1639	Orbital Transfer
1640	Tunnel Diodes
1641	Evaporation
1642	Ionization Of Gases
1643	Atoms
1644	Mos Devices
1645	Formal Logic
1646	Dynamic Response
1647	Fermions
1648	Isotherms
1649	Silicon Nitride
1650	Atmospheric Boundary Layer
1651	Distributed Computer Systems
1652	Ferromagnetic Materials
1653	Thin Film Transistors
1654	Signal Interference
1655	Iron Oxides
1656	Hysteresis
1657	X Ray Diffraction Analysis
1658	Temporal Logic
1659	Concentration (Process)
1660	Molten Materials
1661	Voltage Dividers
1662	Transmissions
1663	Tin Compounds
1664	Atmospheric Movements
1665	Relaxation Processes
1666	Stochastic Models
1667	Activated Carbon Treatment
1668	Microprocessor Chips
1669	Electric Load Forecasting
1670	Dissimilar Metals
1671	Toluene
1672	Insulated Gate Bipolar Transistors (Igbt)
1673	Uncertain Systems
1674	Packet Networks
1675	Mosfet Devices
1676	Model Structures
1677	Impurities
1678	Incompressible Flow
1679	Phase Locked Loops
1680	Conjugated Polymers
1681	Ethers
1682	Optical Instrument Lenses
1683	Monte-Carlo Methods
1684	Three Dimensional Computer Graphics
1685	Metallic Films
1686	Transconductance
1687	Electric Current Control
1688	Temperature Distribution
1689	Space Heating
1690	Numerical Control Systems
1691	Differential Amplifiers
1692	Cost Reduction
1693	Parallel Architectures
1694	Fiber Optic Components
1695	Biochemical Oxygen Demand
1696	Luminescence Of Organic Solids
1697	Fermi Level
1698	Fluid Structure Interaction
1699	Surface Topography
1700	Electrolytic Reduction
1701	Magnetic Flux
1702	Cost Benefit Analysis
1703	Electric Contactors
1704	Cutoff Frequency
1705	X Ray Absorption
1706	Frequency Response
1707	Asymptotic Stability

1708	Control System Stability
1709	Crystalline Materials
1710	Phase Interfaces
1711	Chemicals Removal (Water Treatment)
1712	Surface Tension
1713	Linear Systems
1714	Inlet Flow
1715	Uncertainty Analysis
1716	Shells (Structures)
1717	Switching Networks
1718	Wave Filters
1719	Grain Boundaries
1720	Epitaxial Layers
1721	Cerium Compounds
1722	Particles (Particulate Matter)
1723	Walls (Structural Partitions)
1724	Precipitation (Chemical)
1725	Switching Systems
1726	Boiling Water Reactors
1727	Surface Properties
1728	Adders
1729	Heuristic Algorithms
1730	Phase Separation
1731	Nitrogen Oxides
1732	Coordination Reactions
1733	Conversion Efficiency
1734	Light Absorption
1735	Buffer Layers
1736	Wastewater Reclamation
1737	Nitrogen Removal
1738	Polysilicon
1739	Thermal Noise
1740	Semiconductor Device Manufacture
1741	Electrostatic Discharge
1742	Magnetic Field Measurement
1743	Optimal Systems
1744	Electric Properties
1745	Transistor Transistor Logic Circuits
1746	Analog To Digital Conversion
1747	Robustness (Control Systems)
1748	Crystal Orientation
1749	Field Effect Transistors
1750	Ion Bombardment
1751	Graphite Electrodes
1752	Network Components
1753	Magnetic Devices
1754	Operational Amplifiers
1755	Passivation
1756	Transition Metals
1757	Threshold Voltage
1758	Damping
1759	Measurement Theory
1760	Extrapolation
1761	Wave Propagation
1762	Magnetic Moments
1763	Silicon Compounds
1764	Metal Ions
1765	Heat Conduction
1766	Moisture Determination
1767	Coincidence Circuits
1768	Variable Frequency Oscillators
1769	Gas Permeable Membranes
1770	Surface Structure
1771	Radio Receivers
1772	Titanium Nitride
1773	Semiconductor Doping
1774	Magnetic Field Effects
1775	Surface Active Agents
1776	Microcrystalline Silicon
1777	Two Dimensional
1778	Crystal Atomic Structure
1779	Sewage Pumping Plants
1780	Electron Energy Loss Spectroscopy
1781	Flow Fields
1782	Logic Circuits
1783	Ionization Of Liquids
1784	Cationic Polymerization
1785	Hydrophilicity
1786	Frequency Bands
1787	Pulse Position Modulation
1788	Electric Power Measurement
1789	Drain Current
1790	Surface Roughness
1791	Catalyst Deactivation
1792	Magnetic Storage
1793	Infrared Devices
1794	Vapor Deposition
1795	Flow Patterns
1796	Electric Conductivity
1797	Tungsten Compounds
1798	Free Radical Polymerization
1799	Pressurized Water Reactors
1800	Cyclic Loads
1801	Distribution Functions
1802	Power Spectrum
1803	Electric Lines
1804	Flexible Displays
1805	High Electron Mobility Transistors
1806	Electromagnetic Pulse
1807	Hole Mobility
1808	Residual Stresses
1809	Electric Network Topology
1810	Negative Ions
1811	Logic Gates
1812	Effluent Treatment
1813	Phase Transitions
1814	Low Noise Amplifiers
1815	Two Phase Flow
1816	Rough Set Theory
1817	Electric Signal Systems
1818	Exhaust Gases
1819	Routing Protocols
1820	Continuous Time Systems
1821	Carrier Concentration
1822	Partial Pressure
1823	Terminals (Electric)
1824	Calcination
1825	Trace Elements
1826	Parallel Flow
1827	Chlorine Compounds
1828	Packet Switching

1829	Chemical Vapor Deposition
1830	Sound Insulating Materials
1831	Positive Ions
1832	Phase Noise
1833	Tensile Stress
1834	Bit Error Rate
1835	Voltage Distribution Measurement
1836	Water Cooled Reactors
1837	Telephone Sets
1838	Computer Software Selection And Evaluation
1839	Carrier Lifetime
1840	Waste Heat Utilization
1841	Photoelectrons
1842	Optical Switches
1843	Permittivity
1844	Radial Basis Function Networks
1845	Gateways (Computer Networks)
1846	Gas Mixtures
1847	Fourier Transforms
1848	Logic Devices
1849	Conductive Films
1850	Electric Current Measurement
1851	Grounding Electrodes
1852	Wave Energy Conversion
1853	Plasma Enhanced Chemical Vapor Deposition
1854	Delay Control Systems
1855	Size Separation
1856	Integer Programming
1857	Lyapunov Functions
1858	Two Dimensional Electron Gas
1859	Dielectric Devices
1860	Pulse Code Modulation
1861	Core Disruptive Accidents
1862	Two Term Control Systems
1863	Silicon Oxides
1864	Sliding Mode Control
1865	Three Term Control Systems
1866	Chemical Bonds
1867	Nonlinear Equations
1868	Quadratic Programming
1869	Optical Flows
1870	Ohmic Contacts
1871	Phase Shift
1872	Welded Steel Structures
1873	Transmission Control Protocol
1874	Phase Diagrams
1875	Signal To Noise Ratio
1876	Q Factor Measurement
1877	Electric Field Measurement
1878	Titanium Compounds
1879	Tensile Strain
1880	Electric Breakdown
1881	Nonlinear Distortion
1882	Elementary Particle Sources
1883	Polydispersity
1884	Principal Component Analysis
1885	X Ray Powder Diffraction
1886	Polyimides
1887	Curve Fitting
1888	Natural Frequencies
1889	Fused Salts
1890	Carboxylic Acids
1891	Diffusion In Gases
1892	Linear Matrix Inequalities
1893	Delay Circuits
1894	Combined Sewers
1895	Discrete Time Control Systems
1896	Transformer Windings
1897	Harmonic Distortion
1898	Crystallite Size
1899	Electric Fault Currents
1900	Asymptotic Analysis
1901	Orthogonal Frequency Division Multiplexing
1902	Random Processes
1903	Asynchronous Generators
1904	Energy Dissipation
1905	Digital Signal Processors
1906	Electrostatic Devices
1907	Periodic Structures
1908	Semiconductor Insulator Boundaries
1909	Amides
1910	Rate Constants
1911	Density Functional Theory
1912	Measurement Errors
1913	Flywheel Propulsion
1914	Pulse Width Modulation
1915	Current Voltage Characteristics
1916	Thick Films
1917	Multiplexing Equipment
1918	Inert Gases
1919	Power Supply Circuits
1920	Vector Spaces
1921	Phase Change Memory
1922	Drop Formation
1923	Heat Transfer Coefficients
1924	Equations Of State
1925	Electric Power Factor
1926	Constrained Optimization
1927	Boundary Conditions
1928	Fins (Heat Exchange)
1929	Switching Circuits
1930	Shape Optimization
1931	Bandpass Amplifiers
1932	Timing Devices
1933	Electromagnetic Wave Absorption
1934	Grain Size And Shape
1935	Parameter Extraction
1936	Transmission Line Theory
1937	Parameter Estimation
1938	Timing Circuits
1939	Convex Optimization
1940	Differentiating Circuits
1941	Shape Memory Effect
1942	Cylinders (Shapes)
1943	Voltage Stabilizing Circuits
1944	Approximation Theory
1945	Gates (Transistor)
1946	Wave Equations
1947	Mos Capacitors

1948	Real Variables
1949	Boundary Layers
1950	Equivalent Circuits
1951	Open Circuit Voltage
1952	Proportional Counters
1953	Scattering Parameters
1954	Maximum Likelihood Estimation
1955	Metal Insulator Boundaries
1956	Semiconductor Device Structures
1957	Noise Figure
1958	Optical Band Gaps
1959	Aspect Ratio
1960	Equations Of Motion
1961	Boltzmann Equation
1962	Bond Strength (Chemical)
1963	Shift Registers
1964	Time Varying Systems
1965	Application Specific Integrated Circuits
1966	Hard Disk Storage
1967	Probability Distributions
1968	Bias Voltage
1969	Printed Circuit Manufacture
1970	Time Varying Networks
1971	Polyethers
1972	Hysteresis Loops
1973	Disks (Structural Components)
1974	Convergence Of Numerical Methods
1975	Carrier Sense Multiple Access
1976	Proportional Control Systems
1977	Covariance Matrix
1978	Threshold Logic
1979	Paraffin Waxes
1980	Lagrange Multipliers
1981	Born Approximation
1982	Random Access Storage
1983	Probability Density Function
1984	Saturation Magnetization
1985	Acrylic Monomers
1986	Static Random Access Storage
1987	Integrated Circuit Manufacture
1988	Electric Power Supplies To Apparatus
1989	Flip Flop Circuits
1990	Zero Voltage Switching
1991	Polyethylene Oxides
1992	Polyethylene Terephthalates
1993	Glycols
1994	Quadrature Amplitude Modulation
1995	Lattice Mismatch
1996	Telluric Prospecting
1997	Distributed Parameter Networks
1998	Polyvinyl Chlorides
1999	Frequency Dividing Circuits
2000	Polyethylene Glycols

APPENDIX 7 - COMPLETE RANKED TERM LIST FOR GEOTHERMAL CASE STUDY

1	Myoelectrically Controlled Prosthetics
2	Refractometers
3	Magnetrons
4	Fragrances
5	Phototransistors
6	Nanofibers
7	Nanofluidics
8	Quartz Crystal Microbalances
9	Abrasives
10	Aneroid Altimeters
11	Electron Microscopes
12	Navigation
13	Microscopes
14	Observatories
15	Plastic Molds
16	Steel Metallography
17	Tsunamis
18	Adsorbents
19	Tanning
20	Navigation Systems
21	Nanotechnology
22	Cams
23	Photonics
24	Radiology
25	Diagnostic Radiography
26	Photography
27	Oncology
28	Porcelain
29	Textiles
30	Hidden Markov Models
31	Spacecraft Propulsion
32	Robotics
33	Cytology
34	Broadband Networks
35	Platinum
36	Prosthetics
37	Implants (Surgical)
38	Optical Projectors
39	Dosimetry
40	Feedback
41	Climatology
42	Wind Turbines
43	Medical Imaging
44	Nanowires
45	Positron Emission Tomography
46	Zircon
47	Museums
48	Multimedia Services
49	Recreation Centers
50	Mergers And Acquisitions
51	Electron Probe Microanalysis
52	Precious Metals
53	Patient Rehabilitation
54	Atmospherics
55	Biodiversity
56	Biometrics
57	Radiography
58	Robots
59	Weather Information Services
60	Bacteriology
61	Color Photography
62	Cameras
63	Nuclear Medicine
64	Glass Ceramics
65	Radar
66	Biomineralization
67	Calorimeters
68	Surgery
69	Bioremediation
70	Microbiology
71	Fallout
72	Education
73	Radiotherapy
74	Mice (Computer Peripherals)
75	Photonic Crystal Fibers
76	Pumps
77	Housing
78	Nanocrystalline Alloys
79	Metallurgy
80	Satellite Imagery
81	Tumors
82	Magnets
83	Sun
84	Furnaces
85	Intermetallics
86	Geomagnetism
87	Biogeochemistry
88	Mobile Computing
89	Climate Change
90	Nanocomposites
91	Linguistics
92	Cosmology
93	Crack Tips
94	Haptic Interfaces
95	Bioreactors
96	Reefs
97	Diamond Deposits
98	Plants (Botany)

99	Tobacco
100	Military Photography
101	Petrology
102	Insulin
103	Biomechanics
104	Carbon Nanotubes
105	Innovation
106	Lathes
107	Mems
108	Power Transformers
109	Controllers
110	Biochemistry
111	Fossil Fuels
112	Earth (Planet)
113	Nanotubes
114	Sol-Gels
115	Nitric Oxide
116	Stars
117	Solar Cells
118	Teaching
119	Radon
120	Diagnostic Products
121	Acoustics
122	Nanostructured Materials
123	Network Security
124	Drought
125	Seismology
126	Multimedia Systems
127	Minerals
128	geothermal
129	Research
130	Mining
131	Watermarking
132	Zirconia
133	Genes
134	Plastics
135	Fabrics
136	Superalloys
137	Volcanoes
138	Phosphorylation
139	Quantum Electronics
140	Earthquakes
141	Lasers
142	Radiometry
143	Gadolinium
144	Digital Watermarking
145	Nanocrystals
146	Oceanography
147	Engineering Geology
148	Silver
149	Muscle
150	Repair
151	Tomography
152	Signaling
153	Escherichia Coli
154	Information Management
155	Power Converters
156	Astrophysics
157	Welding
158	Charcoal
159	Nanoparticles
160	Ontology
161	Robot Programming
162	Mineralogy
163	Internet
164	Signal Transduction
165	Earth Sciences
166	Photographic Equipment
167	Reentry
168	Metabolism
169	Forestry
170	Polyaniline
171	Gas Welding
172	Plastic Products
173	Spark Plasma Sintering
174	Turbines
175	Wireless Sensor Networks
176	Parks
177	Gold
178	Magnetic Resonance Imaging
179	Virtual Reality
180	Osmosis
181	Antibiotics
182	Ultrasonic Testing
183	Safety Engineering
184	Tribology
185	Nanorods
186	Cloning
187	Geophysics
188	Ecology
189	Hip Prostheses
190	Architecture
191	Bioactivity
192	Organic Conductors
193	Stainless Steel
194	Transcription
195	Imaging Systems
196	Cell Death
197	Railroads
198	Mobile Devices
199	Bone
200	Water Resources
201	Fatty Acids
202	Wireless Networks
203	Quantum Optics
204	Hydraulics
205	Bayesian Networks
206	Inductively Coupled Plasma
207	Target Drones
208	Rock Products

209	Lighting
210	Pelletizing
211	Dc Motors
212	Shot Peening
213	Granite
214	Interferometry
215	Coagulation
216	Ultrasonography
217	Nanoclusters
218	Statistics
219	Ocean Engineering
220	Probes
221	Batch Reactors
222	Skin
223	Photocatalysts
224	Security Systems
225	Dendrimers
226	Fuels
227	Coatings
228	Microspheres
229	Organometallics
230	Aluminum Castings
231	Glaciers
232	Visualization
233	Dyes
234	Cell Adhesion
235	Polymer Blends
236	Peptides
237	Diamond Cutting Tools
238	Macrophages
239	Software Design
240	Agriculture
241	Atomic Force Microscopy
242	Motors
243	Windows
244	Gas Industry
245	Aerial Photography
246	Benchmarking
247	Synchrotrons
248	Design
249	Authentication
250	Mass Spectrometry
251	Fillers
252	Dna
253	Organic Minerals
254	Cathodic Protection
255	Hydrogeology
256	Surveying
257	Integrated Optoelectronics
258	Engines
259	Rna
260	Gold Deposits
261	Wetlands
262	Textile Industry
263	Hydrophobicity
264	Electrochemistry
265	Geology
266	X Ray Cameras
267	Tides
268	Ecosystems
269	Solutions
270	Lakes
271	Marine Pollution
272	Conservation
273	Pesticides
274	Water Content
275	Fatigue Testing
276	Electrophoresis
277	Supply Chain Management
278	Learning Systems
279	Gas Emissions
280	Buildings
281	Fluidized Beds
282	Smoke
283	Polarimeters
284	Diamonds
285	Computer Graphics
286	Cell Proliferation
287	Data Mining
288	Nuclear Magnetic Resonance
289	Surveys
290	Research Laboratories
291	Hydrology
292	Spacecraft Instruments
293	Computerized Tomography
294	Simulators
295	Coal
296	Spectrometry
297	Standards
298	Optical Filters
299	Nanosheets
300	Magnetohydrodynamics
301	Osteoblasts
302	Radar Imaging
303	Risk Management
304	Coastal Engineering
305	Rapid Prototyping
306	Offshore Petroleum Prospecting
307	Snow
308	Systems Engineering
309	Optical Instruments
310	Blood Vessels
311	Water Management
312	Steel
313	Heat Exchangers
314	Equipment
315	Quality Of Service
316	Transcription Factors
317	Galaxies
318	Geochemistry

319	Computer Software
320	Radar Systems
321	Multiobjective Optimization
322	Sieves
323	Color
324	Ice
325	Steel Research
326	Silanes
327	Metal Detectors
328	Nucleic Acids
329	Nasa
330	Precast Concrete
331	Interoperability
332	Nanospheres
333	Electric Circuit Breakers
334	Glucose
335	Fluorescence
336	Cell Culture
337	Photocatalysis
338	Plant Management
339	Remote Sensing
340	Proteins
341	Physics
342	Crystallography
343	Particle Spectrometers
344	Software Architecture
345	Plasma Diagnostics
346	Optical Microscopy
347	Drilling
348	Testing
349	Architectural Design
350	Dosimeters
351	Polishing
352	Bacteria
353	Petroleum Geology
354	Wastewater
355	Nanocrystalline Materials
356	Petroleum Industry
357	Photonic Crystals
358	Technetium
359	Disinfection
360	Tissue Engineering
361	Superconductivity
362	Transmission Electron Microscopy
363	Conductive Plastics
364	Nonlinear Optics
365	Launching
366	Tin
367	Plant Shutdowns
368	Geochronology
369	Quantum Interference Devices
370	Nitrates
371	Robot Applications
372	Solar Energy
373	Solar Heating
374	Space Optics
375	Wireless Mesh Networks (Wmn)
376	Scaffolds
377	Quality Assurance
378	Towers
379	Iron
380	Friction Stir Welding
381	Remediation
382	Calcium
383	Scanning
384	Spectrometers
385	Intelligent Robots
386	Speckle
387	Ultrasonics
388	Sedimentology
389	Mammals
390	Wells
391	Magnetic Resonance
392	Food Additives
393	Composite Micromechanics
394	Air
395	Photovoltaic Cells
396	Petroleum Reservoir Engineering
397	Sensors
398	Lime
399	Lubricants
400	Enzyme Activity
401	Rivers
402	Image Classification
403	Nuclear Energy
404	Nitration
405	Enzymes
406	Communication
407	Microphase Separation
408	Powder Metallurgy
409	Information Systems
410	Ultrasonic Imaging
411	Joint Prostheses
412	Mass Spectrometers
413	Clouds
414	Diesel Fuels
415	Materials
416	Infrared Detectors
417	Detectors
418	Mesoporous Materials
419	Scanning Electron Microscopy
420	Glycoproteins
421	Radio Altimeters
422	Magnetism
423	Electron Optics
424	Computer Hardware
425	Viscoplasticity
426	Jets

427	Polymers
428	Clay
429	Ultrasonic Applications
430	Computer Programming Languages
431	Electric Motors
432	Gels
433	Recovery
434	Ferrite
435	Multiphoton Processes
436	Strain
437	Computer Simulation Languages
438	Restoration
439	Image Quality
440	Transparency
441	Gasoline
442	Plasmas
443	Herbicides
444	Nanofiltration
445	Fracture
446	Forecasting
447	Water
448	Storms
449	Acids
450	Copper
451	Aircraft
452	Natural Resources Management
453	Tracking Radar
454	Gears
455	Coal Industry
456	Lubrication
457	Wind Power
458	Mechanics
459	Computer Programming
460	Ammonia
461	Nuclear Reactors
462	Pollution
463	Steel Construction
464	Hydroxyapatite
465	Rheology
466	Pathogens
467	Wire
468	Mines
469	Vehicles
470	Coal Gas
471	Plasticity
472	Particle Accelerators
473	Ethanol
474	Pixels
475	Image Analysis
476	Ultra-Wideband (Uwb)
477	Melamine Formaldehyde Resins
478	Rock Mechanics
479	Amino Acids
480	Satellites
481	Chitin
482	Program Interpreters
483	Mortar
484	Tankers (Ships)
485	Fracture Mechanics
486	Ferroelectricity
487	Seed
488	Supercritical Fluids
489	Chitosan
490	Weldability
491	Activation Analysis
492	Adhesion
493	Tectonics
494	Joints (Anatomy)
495	Rubber
496	Radiation
497	Abrasion
498	Molecular Dynamics
499	Toxicity
500	Microwaves
501	Shale Oil
502	Oil Shale
503	Monolayers
504	Targets
505	Computer Systems Programming
506	Laser Ablation
507	Chemical Sensors
508	Glass
509	Monitoring
510	Specifications
511	High Performance Liquid Chromatography
512	Neodymium
513	Natural Gas
514	Fractography
515	Titanium
516	Explosions
517	Electromagnetism
518	Land Use
519	Fruits
520	Reclamation
521	Ellipsometry
522	Raman Spectroscopy
523	Electrospray Ionization
524	Water Pollution
525	Thermoelectric Equipment
526	Mapping
527	Metals
528	Genetic Algorithms
529	Nanocapsules
530	Crops
531	Drives
532	Resins
533	Image Sensors
534	Ac Motors

535	Hydrogels
536	Visual Communication
537	Well Stimulation
538	Satellite Observatories
539	Electric Inverters
540	Oil Fields
541	Thermoplastics
542	Floors
543	Water Conservation
544	Water Quality
545	Optical Radar
546	X Ray Radiography
547	Capacitors
548	Strontium
549	Formaldehyde
550	Powders
551	Java Programming Language
552	Binding Sites
553	Optical Sensors
554	Control
555	Deposits
556	Software Reliability
557	Cell Membranes
558	Hydrogen Peroxide
559	Cobalt
560	Pressure Vessels
561	Fluid Dynamics
562	Rails
563	Turbidity
564	Biochemical Engineering
565	Plasma Interactions
566	Radar Measurement
567	Ion Exchangers
568	Recreational Facilities
569	Interactive Computer Systems
570	Resonance
571	Water Filtration
572	Lithium
573	Laser Applications
574	Phytoplankton
575	Energy Efficiency
576	Ligands
577	Electron Diffraction
578	Catalysis
579	Regulatory Compliance
580	Defects
581	Size Exclusion Chromatography
582	Global System For Mobile Communications
583	Catalysts
584	Motor Transportation
585	Diesel Engines
586	Neural Networks
587	Zinc
588	Alarm Systems
589	Wireless Telecommunication Systems
590	Synthetic Aperture Radar
591	Lead
592	Network Performance
593	Crystal Whiskers
594	Fibers
595	Risk Analysis
596	Semiconductor Quantum Dots
597	Antennas
598	X Ray Diffraction Analysis
599	Geological Surveys
600	Glycerol
601	Solar Power Generation
602	Biodegradation
603	Photoacoustic Spectroscopy
604	Image Retrieval
605	Sodium
606	Cadmium
607	Reservoirs (Water)
608	Biomass
609	Open Systems
610	Processing
611	Biological Materials
612	Calorimetry
613	Query Languages
614	Machine Design
615	Coal Mines
616	Speech Communication
617	Chlorophyll
618	Sugars
619	Radioactivity
620	Couplings
621	Oil Wells
622	Nuclear Magnetic Resonance Spectroscopy
623	Fuzzy Logic
624	Explosives
625	Potassium
626	Aerodynamics
627	Rayon
628	Arsenic
629	Charge Transfer
630	Reinforced Plastics
631	Corrosion
632	Gases
633	Access Control
634	Light
635	Harvesting
636	Optimization
637	Building Materials
638	Quality Control
639	Augmented Reality
640	Well Drilling

641	Oil Sands
642	Phosphorus
643	Rocks
644	Irrigation
645	Water Treatment
646	Speed
647	Image Coding
648	Phospholipids
649	Coal Ash
650	Extrusion
651	Scintillation
652	Film Growth
653	Milling (Machining)
654	Energy Balance
655	Brazing
656	Laws And Legislation
657	Dynamics
658	Thermoanalysis
659	Dissociation
660	Pollution Control
661	Cooling Systems
662	Optical Materials
663	Photosynthesis
664	Morphology
665	Roofs
666	Curing
667	Linear Accelerators
668	Models
669	Fluids
670	Asynchronous Generators
671	Acoustic Noise
672	Ceramic Materials
673	Timber
674	Platinum Compounds
675	Petroleum Transportation
676	Sand Consolidation
677	Thermogravimetric Analysis
678	Salvaging
679	Hydrates
680	Communication Systems
681	Relaxation Time
682	Isotopes
683	Wastewater Treatment
684	Engine Cylinders
685	Earth Atmosphere
686	Cellulose
687	Weather Forecasting
688	Distillation
689	Digital Image Storage
690	Denitrification
691	Integration
692	Extraction
693	Gallium
694	Magnesium
695	Geographic Information Systems
696	Palladium
697	Silicon
698	Association Rules
699	Evolutionary Algorithms
700	Soil Mechanics
701	Carbon Dioxide
702	Particle Detectors
703	Contact Angle
704	Ionizing Radiation
705	Nickel
706	Synchrotron Radiation
707	Biological Systems
708	Ytterbium
709	Ocean Habitats
710	Magnetic Materials
711	Aluminum
712	Toughening
713	Groundwater
714	Mobile Ad Hoc Networks
715	X Ray Analysis
716	Refining
717	Groundwater Resources
718	Strain Energy
719	Zeta Potential
720	Molecular Spectroscopy
721	Composite Coatings
722	Fires
723	Petroleum Refineries
724	Spacecraft
725	Opacity
726	Ceramic Coatings
727	Solvent Extraction
728	Cellular Telephone Systems
729	Focusing
730	Program Compilers
731	Alignment
732	Developing Countries
733	Modulation
734	Chromium
735	Stiffness
736	Biological Membranes
737	Pipe
738	Quantum Theory
739	Image Processing
740	Classifiers
741	Carbon Fiber Reinforced Plastics
742	Speech Processing
743	Nanocrystalline Powders
744	Gas Chromatography
745	Resonators
746	Papermaking
747	Plasma Applications
748	Speech Analysis

749	Cryptography
750	Bone Cement
751	Milling Machines
752	Methane
753	Lanthanum
754	Viscoelasticity
755	Speech Transmission
756	Synthesis (Chemical)
757	Water Supply
758	Liquids
759	Carbon Steel
760	Computer Simulation
761	Water Recycling
762	Textile Processing
763	Ferroelectric Films
764	Life Cycle
765	Data Transfer
766	Multilayer Neural Networks
767	Optical Communication
768	Parallel Architectures
769	Calcium Phosphate
770	Thermodynamics
771	Maleic Anhydride
772	Image Segmentation
773	Metallic Glass
774	Water Cooling Systems
775	Combustion Synthesis
776	Silicon Sensors
777	Corrosion Fatigue
778	Urea
779	Risk Assessment
780	Apatite
781	Blending
782	Soils
783	Computer Operating Systems
784	Energy Transfer
785	Carbon Monoxide
786	Offshore Oil Wells
787	X Ray Diffraction
788	Optoelectronic Devices
789	Semiconductor Growth
790	Batch Cell Culture
791	Barium
792	Drag
793	Molecules
794	Systems Analysis
795	Computer Aided Analysis
796	Fatigue Of Materials
797	Waves
798	Real Time Systems
799	Semantics
800	Mechanical Testing
801	Geomorphology
802	Rhenium
803	Ball Mills
804	Stratigraphy
805	Atmospheric Chemistry
806	Sea Level
807	Natural Gas Well Drilling
808	Environmental Impact
809	Silver Deposits
810	Luminescence
811	Plasma Theory
812	Analytical Geochemistry
813	Reliability
814	Hydrodynamics
815	Embedded Systems
816	Oxygen
817	Water Analysis
818	Chromatographic Analysis
819	Microorganisms
820	Computational Fluid Dynamics
821	Drops
822	Chemical Analysis
823	Real Time Control
824	Infrared Devices
825	Stress Analysis
826	Inductively Coupled Plasma Mass Spectrometry
827	Soil Pollution
828	Diffusers (Optical)
829	Ternary Systems
830	Computational Methods
831	Underwater Structures
832	Scanning Tunneling Microscopy
833	Chemical Attack
834	Land Fill
835	Organic Chemicals
836	Telecommunication Networks
837	Clustering Algorithms
838	Aluminum Metallurgy
839	Fault Detection
840	Diffraction
841	Cavity Resonators
842	Heating
843	Polarographic Analysis
844	Traffic Control
845	Population Statistics
846	Decision Making
847	Organic Acids
848	Radioactive Waste Disposal
849	Fracture Testing
850	Fluidization
851	Synchronous Motors
852	Dust
853	Active Filters
854	Manganese
855	Calcium Chloride
856	Refractory Materials

857	Erosion
858	Errors
859	Field Emission Microscopes
860	Differential Equations
861	Molding
862	Piezoelectric Transducers
863	Substrates
864	Auger Electron Spectroscopy
865	Data Processing
866	Cluster Analysis
867	Cooling
868	Electric Generators
869	Vegetation
870	Remote Control
871	Algorithms
872	Motion Control
873	Superconducting Films
874	Sand
875	Mosfet Devices
876	Plasma Sources
877	Fabrication
878	Activation Energy
879	Speech Recognition
880	Optical Systems
881	Diamond Films
882	Heat Treatment
883	Sulfur
884	Stabilization
885	Photovoltaic Effects
886	Well Logging
887	Rain
888	Petrography
889	Vibration Control
890	Calibration
891	Alloys
892	High Energy Physics
893	Optical Properties
894	Interfaces (Computer)
895	Sandstone
896	Erbium
897	Chemical Activation
898	Structural Geology
899	Heat Transfer
900	Meteorological Radar
901	Near Infrared Spectroscopy
902	Infrared Spectroscopy
903	X Ray Photoelectron Spectroscopy
904	Torque
905	Hydrothermal Synthesis
906	Optical Design
907	Rare Earths
908	Freezing
909	Sols
910	Cooling Water
911	Hydration
912	Molecular Beams
913	Bioactive Glass
914	Magnesium Castings
915	Crosslinking
916	X Rays
917	Electronic Properties
918	Hydrogen
919	Sewage
920	Oil Field Development
921	Copolymers
922	Polymer Matrix Composites
923	Particle Beams
924	Packet Networks
925	Oxygen Vacancies
926	Atomic Spectroscopy
927	Mica
928	Radar Antennas
929	Surface Chemistry
930	Crystals
931	Molybdenum
932	Photoelectricity
933	Ocean Currents
934	Iron Alloys
935	Optical Fibers
936	Scandium
937	Removal
938	Nickel Cadmium Batteries
939	Radioisotopes
940	Diodes
941	Plasma Turbulence
942	Laser Heating
943	Body Fluids
944	Lactic Acid
945	Petroleum Refining
946	Silicon Detectors
947	Shale
948	Image Enhancement
949	Nuclear Propulsion
950	Radiation Detectors
951	Polycyclic Aromatic Hydrocarbons
952	Radiation Effects
953	Solid Solutions
954	Phenolic Resins
955	Atmospheric Aerosols
956	Irradiation
957	Petroleum Analysis
958	Soil Conservation
959	Sodium Sulfate
960	Natural Gas Wells
961	Glass Fibers
962	Fourier Transform Infrared Spectroscopy
963	Image Reconstruction
964	Nitrogen

965	Plasma Waves
966	Heavy Metals
967	Telecommunication Equipment
968	Magnetic Devices
969	Carbon Films
970	Graphite
971	Textures
972	Weathering
973	Inertial Confinement Fusion
974	Water Injection
975	Matlab
976	Ph Effects
977	Hardening
978	Experiments
979	Creep
980	Functional Polymers
981	Semiconductor Quantum Wires
982	Cements
983	Object Oriented Programming
984	Hydrocarbons
985	Ordinary Differential Equations
986	Fuel Oils
987	Geologic Models
988	Electric Instrument Transformers
989	Magnetic Fields
990	Thermoelectricity
991	Green'S Function
992	Martensitic Transformations
993	Diffusion
994	Organic Carbon
995	Dc Generators
996	Stability
997	Doppler Radar
998	Solar Radiation
999	Photodegradation
1000	Radiometers
1001	Data Compression
1002	Plasma Jets
1003	Cracks
1004	Seismic Design
1005	Pressure Effects
1006	Sodium Chloride
1007	Lignin
1008	Phosphates
1009	Plasma Torches
1010	Electric Power Generation
1011	Geothermal Springs
1012	Sensor Networks
1013	Activated Carbon
1014	River Pollution
1015	Mercury (Metal)
1016	Titanium Dioxide
1017	Structural Design
1018	Hardness
1019	Graphical User Interfaces
1020	Contamination
1021	Binding Energy
1022	Mathematical Morphology
1023	Petroleum Prospecting
1024	Crystal Growth
1025	Compression Testing
1026	Runoff
1027	Microelectromechanical Devices
1028	Mineral Oils
1029	Tellurium Compounds
1030	Acoustic Fields
1031	Soil Moisture
1032	Refractive Index
1033	Heterogeneous Networks
1034	Fuzzy Control
1035	Process Control
1036	Solid Oxide Fuel Cells (Sofc)
1037	Pore Size
1038	Adaptive Control Systems
1039	Thin Films
1040	Temperature
1041	Catalytic Oxidation
1042	Film Preparation
1043	Europium
1044	Specific Heat
1045	Vacuum
1046	Local Area Networks
1047	Moisture Control
1048	Raman Scattering
1049	Temperature Control
1050	Materials Handling
1051	Alumina
1052	Germanium
1053	Highway Systems
1054	Spectroscopic Analysis
1055	Acetone
1056	Ionic Liquids
1057	Electric Network Analysis
1058	Porphyrins
1059	Durability
1060	Structural Analysis
1061	Materials Properties
1062	Metal Casting
1063	Hydrogen Storage
1064	Epitaxial Growth
1065	Titanium Alloys
1066	Surface Testing
1067	Bicmos Technology
1068	Signal Detection
1069	Solvents
1070	Mechanical Properties
1071	Propane

1072	Degradation
1073	Degassing
1074	Societies And Institutions
1075	Debris
1076	Gamma Rays
1077	High Resolution Transmission Electron Microscopy
1078	Polymer Films
1079	Sensor Nodes
1080	Hydrogen Storage Alloys
1081	Total Quality Management
1082	Mass Transfer
1083	Colliding Beam Accelerators
1084	Ion Exchange
1085	Garnets
1086	Knowledge Based Systems
1087	Dynamic Programming
1088	Pyroelectricity
1089	Signal Processing
1090	Energy Conversion
1091	Reaction Kinetics
1092	Transient Analysis
1093	Oxidation
1094	Zeolites
1095	Environmental Protection Agency
1096	Methanol
1097	Knowledge Acquisition
1098	Vibration Analysis
1099	Water Levels
1100	Conformations
1101	Solidification
1102	Magnetic Properties
1103	Strain Rate
1104	Multilayers
1105	High Pressure Liquid Chromatography
1106	Emission Spectroscopy
1107	Smelting
1108	Styrene
1109	Differential Scanning Calorimetry
1110	Optical Films
1111	Propellants
1112	Cellular Radio Systems
1113	Plasticizers
1114	Beryllium
1115	Silicon Carbide
1116	Flocculation
1117	Control System Analysis
1118	Boron
1119	Combustion
1120	Induction Motors
1121	Sensitivity Analysis
1122	Helium
1123	Waveform Analysis
1124	Global Positioning System
1125	Acceleration
1126	Hydrogen Production
1127	Inorganic Coatings
1128	Protective Coatings
1129	Neodymium Lasers
1130	Fuel Additives
1131	Nonlinear Analysis
1132	Mathematical Models
1133	Epitaxial Films
1134	Electrodes
1135	Seismic Prospecting
1136	Flow Of Fluids
1137	Thermochemistry
1138	Magnetoresistance
1139	Quartz
1140	Silica
1141	Multivariant Analysis
1142	Interferometers
1143	Electric Automobiles
1144	Pinch Effect
1145	Problem Solving
1146	Charge Trapping
1147	Information Fusion
1148	Silicon Alloys
1149	Internet Protocols
1150	Chemical Reactions
1151	Voltage Control
1152	Electromagnetic Fields
1153	Process Monitoring
1154	Stress Concentration
1155	Tensors
1156	Suspensions (Fluids)
1157	Stoichiometry
1158	Zinc Sulfide
1159	Submarine Geology
1160	Oil Resistance
1161	Cost Benefit Analysis
1162	Feature Extraction
1163	Metal Analysis
1164	Leaching
1165	Grafting (Chemical)
1166	Microstructure
1167	Safety Factor
1168	Concrete Buildings
1169	Dispersions
1170	Reynolds Number
1171	Photoacoustic Effect
1172	Unsaturated Polymers
1173	Graphic Methods
1174	Magnetic Anisotropy
1175	High Pressure Effects
1176	Dynamic Mechanical Analysis
1177	Pulsed Laser Deposition
1178	Metal Extrusion

1179	Alkylation
1180	Flow Simulation
1181	Heavy Water
1182	Serpentine
1183	Surface Analysis
1184	Edge Detection
1185	Nucleation
1186	Structural Dynamics
1187	Decision Support Systems
1188	Network Protocols
1189	Seawater
1190	Reinforcement
1191	Compaction
1192	Electric Drives
1193	Estimation
1194	Structural Optimization
1195	Vanadium
1196	Epoxy Resins
1197	Metal Recovery
1198	Traffic Surveys
1199	Electric Wire
1200	Traffic Congestion
1201	Plasma Deposition
1202	Soldering Alloys
1203	Organic Compounds
1204	Wireless Local Area Networks (Wlan)
1205	Adsorption
1206	Fusion Reactors
1207	Coal Combustion
1208	Wind Stress
1209	Regression Analysis
1210	Estuaries
1211	Crystallization
1212	Routing Algorithms
1213	Zinc Oxide
1214	Water Waves
1215	Deposition
1216	Functional Groups
1217	Chemical Oxygen Demand
1218	Agricultural Chemicals
1219	Plastic Films
1220	Logging (Forestry)
1221	Crystal Structure
1222	Wear Of Materials
1223	Surface Treatment
1224	Neodymium Alloys
1225	Ores
1226	Photoresists
1227	Oxide Minerals
1228	Sampling
1229	Microchannels
1230	Vibrators
1231	Synchronization
1232	Desorption
1233	Reflection
1234	Tungsten
1235	Ferritic Steel
1236	Polyethylenes
1237	Hafnium
1238	Ethylene
1239	Computational Complexity
1240	Organic Polymers
1241	Target Tracking
1242	Permanent Magnets
1243	Fault Tolerant Computer Systems
1244	Photoluminescence
1245	Lanthanum Alloys
1246	Calculations
1247	Reconstruction (Structural)
1248	Temperature Programmed Desorption
1249	Dynamic Analysis
1250	Redox Reactions
1251	Ion Exchange Resins
1252	Melting
1253	Redundancy
1254	Soil Testing
1255	Flue Gases
1256	Liquefaction
1257	Power Transmission
1258	Ignition
1259	Molecular Beam Epitaxy
1260	Structural Properties
1261	Niobium
1262	Field Programmable Gate Arrays (Fpga)
1263	Surface Morphology
1264	Mobile Telecommunication Systems
1265	Thiophene
1266	Structural Frames
1267	Concrete Construction
1268	Crystallization Kinetics
1269	Earthquake Effects
1270	Torque Control
1271	Supply Chains
1272	Annealing
1273	Phase Modulation
1274	Agglomeration
1275	Adsorption Isotherms
1276	Optical Emission Spectroscopy
1277	Toluene
1278	Alkalinity
1279	Pyrolysis
1280	Data Structures
1281	Tungsten Carbide
1282	Cesium
1283	Artificial Limbs
1284	Magnetic Domains
1285	Microwave Irradiation

1286	Benzene
1287	Spectrum Analysis
1288	Failure Analysis
1289	Polarization
1290	Electric Load Forecasting
1291	Spin Dynamics
1292	Ad Hoc Networks
1293	Scattering
1294	Piezoelectric Ceramics
1295	Control Theory
1296	Ferroelectric Ceramics
1297	Light Emission
1298	Acetic Acid
1299	Oxidation Resistance
1300	Petroleum Reservoir Evaluation
1301	Well Completion
1302	Exploratory Geochemistry
1303	Biological Water Treatment
1304	Self Assembly
1305	Absorption
1306	Thin Film Devices
1307	Iron Compounds
1308	Percolation (Fluids)
1309	High Temperature Applications
1310	Ore Deposits
1311	Electric Fields
1312	Gadolinium Alloys
1313	Separation
1314	Electrochemical Impedance Spectroscopy
1315	Metal Refining
1316	Electrochemical Corrosion
1317	Propylene
1318	Esterification
1319	Pressure Drop
1320	Nozzles
1321	Rare Earth Alloys
1322	Thermal Conductivity
1323	Linear Programming
1324	Oil Field Equipment
1325	System Stability
1326	Esters
1327	Titanium Carbide
1328	Magnesium Alloys
1329	Stresses
1330	Electric Properties
1331	Risk Perception
1332	Wind Effects
1333	Dissolved Oxygen
1334	Atmospheric Radiation
1335	Tensile Testing
1336	Lithology
1337	Coal Deposits
1338	Readout Systems
1339	Diamond Like Carbon Films
1340	Structural Metals
1341	Simulated Annealing
1342	Concretes
1343	Magnetic Field Effects
1344	Electrochemical Properties
1345	Seismic Response
1346	Water Vapor
1347	Sulfate Minerals
1348	Monte Carlo Methods
1349	Inclusions
1350	Pore Pressure
1351	Free Radical Reactions
1352	Paraffins
1353	Digital To Analog Conversion
1354	Silicon Wafers
1355	Aldehydes
1356	Aluminum Powder Metallurgy
1357	Laser Pulses
1358	Cerium
1359	Field Emission
1360	Aniline
1361	Groundwater Pollution
1362	Perovskite
1363	Potable Water
1364	Fracture Toughness
1365	Laser Excitation
1366	Gallium Alloys
1367	Neutron Radiography
1368	Numerical Methods
1369	Water Treatment Plants
1370	Surface Measurement
1371	Tubes (Components)
1372	Ultraviolet Spectroscopy
1373	Numerical Analysis
1374	Enhanced Recovery
1375	Gallium Nitride
1376	Amines
1377	Reliability Analysis
1378	Tantalum
1379	Number Theory
1380	Aromatic Hydrocarbons
1381	Surface Waves
1382	Surfaces
1383	Steel Structures
1384	Ocean Structures
1385	Multiwalled Carbon Nanotubes (Mwcn)
1386	Taxonomies
1387	Amorphous Alloys
1388	Spontaneous Emission
1389	Molecular Sieves
1390	Petroleum Reservoirs
1391	Particle Size

1392	Catalyst Activity
1393	Associative Processing
1394	Coupling Agents
1395	Olefins
1396	Hamiltonians
1397	Aromatic Compounds
1398	Zirconium Alloys
1399	Oxide Films
1400	Stochastic Models
1401	Pulsed Laser Applications
1402	Chemical Stability
1403	Scintillation Counters
1404	Particle Interactions
1405	Ketones
1406	Dynamic Models
1407	Chromophores
1408	Support Vector Machines
1409	Rubidium
1410	Atomic Absorption Spectrometry
1411	Piezoelectricity
1412	Fire Hazards
1413	Landforms
1414	Brittle Fracture
1415	Physical Properties
1416	Temperature Measuring Instruments
1417	Paper Sheeting
1418	Harmonic Analysis
1419	Yttrium
1420	Crystal Atomic Structure
1421	Induction Heating
1422	Piezoelectric Materials
1423	Soil Pollution Control
1424	Size Distribution
1425	X Ray Powder Diffraction
1426	Decision Theory
1427	Microcrystalline Silicon
1428	Catalyst Selectivity
1429	Atmospheric Temperature
1430	Tensile Strength
1431	Optimal Systems
1432	Dewatering
1433	Sediments
1434	Grain Growth
1435	Hysteresis
1436	Photons
1437	Slags
1438	Electric Potential
1439	Wetting
1440	Argon
1441	Industrial Waste Treatment
1442	Fault Tolerance
1443	Surface Defects
1444	Silicon Compounds
1445	Gelation
1446	Function Evaluation
1447	Power Spectrum
1448	Complexation
1449	Correlation Methods
1450	Gas Adsorption
1451	Organic Solvents
1452	Viscosity
1453	Evaporation
1454	Polymer Melts
1455	Electromagnetic Waves
1456	Potassium Compounds
1457	Porous Materials
1458	Time Domain Analysis
1459	Semiconducting Zinc Compounds
1460	Silica Gel
1461	Surface Properties
1462	Thermodynamic Properties
1463	Quenching
1464	Sintering
1465	Zirconium
1466	Fused Silica
1467	Plasma Heating
1468	Semiconducting Bismuth Compounds
1469	Ethylene Glycol
1470	Piles
1471	Amorphous Materials
1472	Solid Lubricants
1473	Light Sources
1474	Gas Condensates
1475	Zinc Compounds
1476	Sorption
1477	Metallic Films
1478	Effluents
1479	Three Dimensional
1480	Liquid Crystals
1481	Calcium Alloys
1482	Indium Arsenide
1483	Ionization Of Liquids
1484	Calcination
1485	Infrared Radiation
1486	Press Load Control
1487	Refraction
1488	Yield Stress
1489	Mixtures
1490	Concentration (Process)
1491	Laser Produced Plasmas
1492	Absorption Spectroscopy
1493	Transfer Functions
1494	Multiplexing
1495	Dynamic Response
1496	Martensite
1497	Cerium Alloys
1498	Surface Waters

1499	Atoms
1500	Aquifers
1501	Sedimentation
1502	Phase Diagrams
1503	Statistical Tests
1504	Mechanical Alloying
1505	Sol-Gel Process
1506	Boron Compounds
1507	Amorphous Carbon
1508	Large Eddy Simulation
1509	Thermal Logging
1510	Nickel Alloys
1511	Ferromagnetic Materials
1512	Decision Trees
1513	Discriminant Analysis
1514	Strength Of Materials
1515	Pressure Distribution
1516	Electric Impedance Tomography
1517	Ultraviolet Radiation
1518	Ball Milling
1519	Waste Disposal
1520	Phosphate Minerals
1521	Particle Size Analysis
1522	Semiconducting Gallium
1523	Piezoelectric Actuators
1524	Rock Bursts
1525	Ion Implantation
1526	Crystalline Materials
1527	Cracking (Chemical)
1528	Curve Fitting
1529	Ions
1530	Fly Ash
1531	Hydraulic Fracturing
1532	Face Recognition
1533	Catchments
1534	Cyclic Voltammetry
1535	Magnetron Sputtering
1536	Amorphous Silicon
1537	Neutrons
1538	Water Absorption
1539	Impact Strength
1540	Doping (Additives)
1541	Underground Explosions
1542	Hydrogenation
1543	Monomers
1544	Electric Lines
1545	Conversion Efficiency
1546	Nitrides
1547	Unmanned Aerial Vehicles (Uav)
1548	Optical Flows
1549	Dissolution
1550	Domes
1551	Linear Control Systems
1552	Orthogonal Frequency Division Multiplexing
1553	Two Dimensional
1554	Semiconducting Silicon
1555	Corrosion Resistance
1556	Phosphate Coatings
1557	Polycrystalline Materials
1558	Atmospheric Pressure
1559	Seismic Waves
1560	Bentonite
1561	Electric Power Transmission
1562	Pulse Modulation
1563	Pumping Plants
1564	Electric Power Transmission Networks
1565	Approximation Algorithms
1566	Frequency Response
1567	Cmos Integrated Circuits
1568	Liquid Metals
1569	Melting Point
1570	Impurities
1571	Non Newtonian Flow
1572	Surface Reconstruction
1573	Subsidence
1574	Copper Deposits
1575	Heat Flux
1576	Shielding
1577	Lithium Alloys
1578	Nitrogen Compounds
1579	Radioactive Elements
1580	Amorphous Films
1581	Semiconducting Indium
1582	Conducting Polymers
1583	Stimulated Raman Scattering
1584	Electron Beams
1585	Space Debris
1586	Ostwald Ripening
1587	Copper Alloys
1588	Shear Stress
1589	Crystallite Size
1590	Bubbles (In Fluids)
1591	Vectors
1592	Deformation
1593	Resource Allocation
1594	Velocity
1595	Magnetic Flux
1596	Friction
1597	Neutron Sources
1598	Secondary Batteries
1599	Catalyst Supports
1600	Sulfur Compounds
1601	Trace Analysis
1602	Wafer Bonding
1603	Heavy Oil Production
1604	Polystyrenes
1605	Copolymerization
1606	Three Dimensional Computer Graphics

1607	Silicon Oxides
1608	Iridium
1609	Single Crystals
1610	Sediment Transport
1611	Phenols
1612	Oxides
1613	Hydrochloric Acid
1614	Mechanical Permeability
1615	X Ray Scattering
1616	Temperature Distribution
1617	Amides
1618	Modal Analysis
1619	Network Layers
1620	Thermal Effects
1621	Particle Beam Tracking
1622	Welds
1623	Proton Beams
1624	Superconducting Magnets
1625	Dendrites (Metallography)
1626	Electron Energy Loss Spectroscopy
1627	Industrial Wastes
1628	Technological Forecasting
1629	Backscattering
1630	Surface Active Agents
1631	Metallic Compounds
1632	Ceramic Capacitors
1633	Nickel Oxide
1634	Navier Stokes Equations
1635	Spheres
1636	Fretting Corrosion
1637	Phase Stability
1638	Stiffness Matrix
1639	Flow Patterns
1640	Mechanical Stability
1641	Heptane
1642	Error Analysis
1643	Stress Corrosion Cracking
1644	Two Phase Flow
1645	Azo Dyes
1646	Leakage (Fluid)
1647	Isotherms
1648	Millimeter Waves
1649	Surface Structure
1650	Graph Theory
1651	Two Term Control Systems
1652	Cement Manufacture
1653	Electric Fault Location
1654	Gas Permeability
1655	Melt Spinning
1656	Semiconducting Germanium Compounds
1657	Wastewater Reclamation
1658	Optical Transfer Function
1659	Bit Error Rate
1660	Powder Magnetic Cores
1661	Optical Phase Conjugation
1662	Carboxylic Acids
1663	Elasticity
1664	Hydrophilicity
1665	Ammonium Compounds
1666	Reliability Theory
1667	Set Theory
1668	Electromagnetic Wave Diffraction
1669	Computer Software Selection And Evaluation
1670	Metal Ions
1671	Solid State Reactions
1672	Glass Transition
1673	Transients
1674	X Ray Apparatus
1675	Linearization
1676	Volatile Organic Compounds
1677	Thermodynamic Stability
1678	Agricultural Runoff
1679	Encapsulation
1680	Phase Interfaces
1681	Photolysis
1682	Cerium Compounds
1683	Semiconductor Device Manufacture
1684	Probability
1685	Polycondensation
1686	Rotation
1687	Carbides
1688	Polypropylenes
1689	Triangulation
1690	Basalt
1691	Object Recognition
1692	Manganese Compounds
1693	Plasma Spraying
1694	Igneous Rocks
1695	Orbits
1696	Highway Traffic Control
1697	Clay Minerals
1698	Dielectric Properties
1699	Hydrogen Bonds
1700	Semiconducting Organic Compounds
1701	Stream Flow
1702	Random Number Generation
1703	Residual Fuels
1704	Chemical Vapor Deposition
1705	Precipitation (Chemical)
1706	Reinforced Concrete
1707	Iron Ores
1708	Rigidity
1709	Field Effect Transistors

1710	Sublimation
1711	Light Absorption
1712	Aluminum Alloys
1713	Motion Compensation
1714	Protons
1715	Volcanic Rocks
1716	Surface Diffusion
1717	Roads And Streets
1718	Solid Electrolytes
1719	Silicon On Insulator Technology
1720	Electrons
1721	Air Permeability
1722	Distribution Functions
1723	Viscous Flow
1724	Photomultipliers
1725	Phase Transitions
1726	Nickel Deposits
1727	Titanium Oxides
1728	Wear Resistance
1729	Lubricating Oils
1730	Heterojunction Bipolar Transistors
1731	Coordination Reactions
1732	Cobalt Compounds
1733	Strontium Compounds
1734	Size Determination
1735	Uncertainty Analysis
1736	Negative Ions
1737	Positive Ions
1738	Electric Resistance
1739	Plates (Structural Components)
1740	Arid Regions
1741	Electrochemical Electrodes
1742	Phase Separation
1743	Neutron Emission
1744	Troposphere
1745	Ethers
1746	Rietveld Analysis
1747	Rapid Solidification
1748	Polyethylene Glycols
1749	Velocity Measurement
1750	Groundwater Flow
1751	Sodium Alloys
1752	Sprayed Coatings
1753	Meteorological Instruments
1754	Buoyancy
1755	Random Processes
1756	Austenite
1757	Density Functional Theory
1758	Fading Channels
1759	Salinity Measurement
1760	Zirconium Compounds
1761	Plasma Enhanced Chemical Vapor Deposition
1762	Coastal Zones
1763	Capillarity
1764	Positrons
1765	Approximation Theory
1766	Hexane
1767	Titanium Nitride
1768	Nickel Compounds
1769	Chemicals Removal (Water Treatment)
1770	Yttrium Alloys
1771	Sulfur Deposits
1772	Iron Oxides
1773	Shoulders (Road)
1774	Dielectric Properties Of Solids
1775	Combustion Chambers
1776	Silicon Nitride
1777	Horizontal Wells
1778	Semiconducting Silicon Compounds
1779	Compressive Strength
1780	Vapors
1781	Porosity
1782	Atmospheric Spectra
1783	Coupled Circuits
1784	Electromagnetic Wave Backscattering
1785	Shear Strain
1786	Electric Rectifiers
1787	Hydraulic Structures
1788	Radiative Transfer
1789	Hydrostatic Pressure
1790	Bearings (Structural)
1791	Strengthening (Metal)
1792	Principal Component Analysis
1793	Electric Conductivity
1794	Olivine
1795	Phosphors
1796	Capacitance
1797	Crystal Lattices
1798	Gas Permeable Membranes
1799	Carbonates
1800	Copper Oxides
1801	Effluent Treatment
1802	Frequency Bands
1803	Multicarrier Modulation
1804	Forced Convection
1805	Fatigue Crack Propagation
1806	Stators
1807	Charge Coupled Devices
1808	Silicates
1809	Torsional Stress
1810	Radar Target Recognition
1811	Iron Removal (Water Treatment)
1812	Hole Concentration
1813	Mullite
1814	Nitrogen Oxides
1815	Dielectric Waveguides

1816	Rapid Thermal Annealing
1817	Eutectics
1818	Finite Element Method
1819	Soil Structure Interactions
1820	Geometrical Optics
1821	Magnetite
1822	Vapor Pressure
1823	Radiation Shielding
1824	Chlorine Compounds
1825	Electric Power Factor
1826	Plastic Deformation
1827	Calcite
1828	Solid Wastes
1829	Deposition Rates
1830	Quality Function Deployment
1831	Bending Strength
1832	Surface Tension
1833	Dynamic Recrystallization
1834	Lead Alloys
1835	Crystal Impurities
1836	Earthquake Resistance
1837	Grouting
1838	Shells (Structures)
1839	Nonionic Surfactants
1840	Paraffin Waxes
1841	Routing Protocols
1842	Shear Strength
1843	Surface Roughness
1844	Counting Circuits
1845	Electromagnetic Pulse
1846	Heuristic Algorithms
1847	Transpiration
1848	Evaporative Cooling Systems
1849	Positron Annihilation Spectroscopy
1850	Maintainability
1851	Reluctance Motors
1852	Passivation
1853	Mixed Convection
1854	Excavation
1855	Ductility
1856	Electromagnetic Wave Scattering
1857	Application Programming Interfaces (Api)
1858	Rate Constants
1859	Sedimentary Rocks
1860	Signal To Noise Ratio
1861	Interpolation
1862	Random Variables
1863	Polymethyl Methacrylates
1864	Polyvinyl Alcohols
1865	Radioactive Wastes
1866	Dislocations (Crystals)
1867	Chromium Compounds
1868	Finite Difference Time Domain Method
1869	Aspect Ratio
1870	Trace Elements
1871	Damping
1872	Fourier Transforms
1873	Transmission Line Theory
1874	Optical Resolving Power
1875	Pulse Width Modulation
1876	Seepage
1877	Sulfide Minerals
1878	Elastic Waves
1879	Inverse Problems
1880	Parameter Estimation
1881	Polyethylene Oxides
1882	Soil Surveys
1883	Wave Equations
1884	Negative Temperature Coefficient
1885	Atmospheric Movements
1886	Natural Gas Well Completion
1887	Bearing Capacity
1888	Elastoplasticity
1889	Density (Specific Gravity)
1890	Cadmium Compounds
1891	Amination
1892	Ion Bombardment
1893	Permittivity
1894	Carbonate Minerals
1895	Aerodynamic Loads
1896	Equations Of Motion
1897	Polyvinyl Chlorides
1898	Rain Gages
1899	Semiconducting Lead Compounds
1900	Well Spacing
1901	Binary Codes
1902	Open Circuit Voltage
1903	Convolution
1904	Rare Earth Elements
1905	Bending (Deformation)
1906	Corundum
1907	Electric Impedance
1908	Backpropagation
1909	Conjugate Gradient Method
1910	Probability Distributions
1911	Residual Stresses
1912	Internal Friction
1913	Biochemical Oxygen Demand
1914	Frequency Allocation
1915	Synchronous Generators
1916	Electromagnetic Wave Emission
1917	Crack Initiation
1918	Electric Discharges

1919	Surface Topography
1920	Method Of Moments
1921	Moisture Determination
1922	Heat Affected Zone
1923	Percolation (Solid State)
1924	Shear Waves
1925	Silver Alloys
1926	Shear Bands
1927	Boltzmann Equation
1928	Permittivity Measurement
1929	Fracturing Fluids
1930	Boundary Layers
1931	Remanence
1932	Energy Dissipation
1933	Integer Programming
1934	Guided Electromagnetic Wave Propagation
1935	Radar Cross Section
1936	Nanocantilevers
1937	Brittleness
1938	Proportional Control Systems
1939	Communication Channels (Information Theory)
1940	Electromagnetic Wave Polarization
1941	Flammability
1942	Overpasses
1943	Temperature Indicating Cameras
1944	Silicate Minerals
1945	Elastic Moduli
1946	Inductance
1947	Turbogenerators
1948	Convergence Of Numerical Methods
1949	Percolation (Computer Storage)
1950	Electric Network Topology
1951	Grain Size And Shape
1952	Kaolinite
1953	Synthetic Apertures
1954	Feldspar
1955	Probability Density Function
1956	Grain Boundaries
1957	Thermionic Emission
1958	Refuse Disposal
1959	Slip Forming
1960	Molecular Orbitals
1961	Crystalline Rocks
1962	Metal Vapor Lamps
1963	Radioactive Prospecting
1964	Alloying
1965	Metallorganic Chemical Vapor Deposition
1966	Leachate Treatment
1967	Spontaneous Potential Logging
1968	Rayleigh Fading
1969	Pyrites
1970	Transformer Windings
1971	Compression Ratio (Machinery)
1972	Ultrahigh Molecular Weight Polyethylenes
1973	Covariance Matrix
1974	Grain Refinement
1975	Geodetic Satellites
1976	Ore Deposit Geology
1977	Settling Tanks
1978	Tantalate Minerals
1979	Electric Fault Currents
1980	Refuse Incinerators
1981	Pollution Induced Corrosion
1982	Directional Patterns (Antenna)
1983	Magnesium Printing Plates
1984	Pile Foundations
1985	Nonmetallic Matrix Composites
1986	Grain Boundary Sliding
1987	Bombs (Ordnance)
1988	Eigenvalues And Eigenfunctions
1989	Low Permeability Reservoirs
1990	Weibull Distribution
1991	Fins (Heat Exchange)
1992	Rock Bolting
1993	Least Squares Approximations
1994	Radioactivity Logging
1995	Railroad Plant And Structures
1996	Clay Alteration
1997	Moire Fringes
1998	Alloying Elements
1999	Recharging (Underground Waters)
2000	Single-Walled Carbon Nanotubes (Swcn)