Winter 12-5-2004

Maximising Technology Efficiencies for SMEs Using Computer Intelligence

Adrian Collins

Maximising Technology Efficiencies for SMEs Using Computer Intelligence

Adrian Collins

University of Wollongong, Australia
adrian@uow.edu.au

ABSTRACT

This paper is part of a series which present elements of our research into technology adoption and usability problems faced by small businesses (SME's) in Australia. We discuss our approach to research in the small business area, the importance of aggregating small businesses into vertical technology sectors and the usability improvements which can be expected from this collaborative approach. In this paper we focus on information retrieval requirements and the related usability problems which confront small businesses. We discuss adoption and usability issues, requirements and risks using a business case study, then introduce experiments which it is hoped will deliver efficiency improvements to information retrieval applications for small businesses.

Keywords; Small Business, collaborative approach, content ranking

1. INTRODUCTION

Our goal is to help SME's find innovative and novel ways to utilise technology. We believe SME's can significantly benefit from a selective adoption of technologies to improve efficiencies for staff by improving communication and information sharing facilities with other organisations. Technology adoption can be extremely cost effective for small business and provide a genuine return on short term investment. Our research aim is to test the viability of improving access to technology for SME's by using a collaborative approach to investment, acquisition and adoption of information and communication technology. It is hoped that this approach will allow individual SME's to spread the cost and risk of eBusiness adoption and use by entering into vertical market collectives. Typically, around 25% of Australian businesses are making some use of electronic business (eBusiness) to reduce the cost of doing business or improve profit potential. Recent research into the local region has found that eBusiness adoption by SMEs is as low as 6%. The major reasons for not adopting eBusiness technology by SMEs:

- Not relevant to the business profile
- Technology is too complicated
- The organisation does not have the resources to implement the technology

2. AGGREGATING SME'S INTO A VERTICAL SECTOR

In Australia, many small businesses are focused on supporting a single, larger business entity or market area. In this case, the SME will have implemented the necessary technology and infrastructure to complete the business process in the most profitable way but lacks the resources to further expand. Many organisations would like to grow their business but due to a lack of resources (finances, staff, expertise etc.) cannot take advantage of growth opportunities. This lack of resources also acts as a barrier to adopting technology which allows them to expand and service new markets. We propose a collaborative approach in which small business stakeholders can pool resources and construct an eBusiness facility that all share. This approach fosters closer business relationships and allows individual SMEs to pool resources by aligning organisations with similar attributes into vertical markets (technology wise). Once organisations with similar attributes are aligned, benefits can be derived by developing and implementing common software solutions which both reduce the cost of doing business and maximise operational efficiencies within the organisation.

An example of organisations with similar attributes is the mercantile agenct market, which has more than 1000 small, independent organisations operating in Australia. A mercantile agent acts on a client's behalf and represents persons and firms by collecting overdue debt and by identifying individuals and organisations implicated in credit fraud and theft. When these types of small organisations are analyzed, we observe that while they are physically separate small organisations involved in a specific business task, the area and scope of their business overlaps that of other small businesses involved in similar operational activities. In the main, mercantile agencies perform services such as:

- Collection: make demands for payment of overdue debts
- Investigation: obtain information and report on the personal character, occupation, activities or actions of persons or businesses
- Process serving: deliver court issued documents for a variety of reasons including debt recovery, family law proceedings, civil and criminal litigation.
- Repossession: act for principals when financial contracts (leases, bills of sale, hire purchase agreements etc.) have been broken. Seeking payment of arrears or the balance of a loan facility

or else to recover the goods on behalf of the principal.

While the operational activities are different among the organisations within the vertical market, there is a significant overlap in many areas. Agents share the same types of clients, acting for banks, financiers, lawyers, insurers, the general business community or individuals. In addition, they share many similar information processing requirements. Document Management Systems (DMS), precedents, document archival, mail merging, diary systems, backup and disaster recovery.

We hypothesize that small businesses can save considerable capital investment dollars by aggregating their processing requirements and benefit by collaboration [1]. These benefits are achieved via the use of portal technology which has been shown to result in a number of business and technical benefits:

- Increased productivity and efficiency
- Provides organisations with the ability to present a central point of access to data and organizational documentation
- Facilitates mobile and pervasive device access
- Increases marketing opportunities
- Provides access to component based solutions

3. CASE STUDY

A mercantile agency typically follows the business model shown in figure 1. Currently 10% of matters are handled outside the local court domain. The business sees this as a major growth area, but the trade off is being able to maximise business efficiencies and promote a larger client base against maintaining a small number of personnel. In order to maximise business efficiencies, the organisation invests in technology, deciding that technology adoption will grow the business exponentially, while investment in staff with only grow the business according to the number of additional staff members.

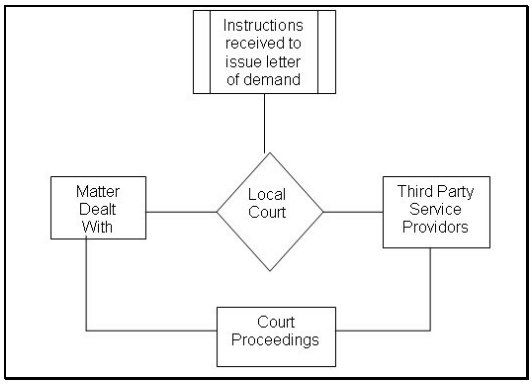

Figure 1.

When the matter is to be dealt with (handled locally), the staff communicate with the debtor and local court on the clients behalf, this process may also include communication with solicitors on both sides. However, when the debtor resides in another locality, the matter may need to be heard in that court locality. In this case, service providers who may act on the mercantile agencies behalf must be located, fees negotiated, and paid before action can commence in that locality.

This presents a major problem for the organisation, because there is no recognised registry of agents. The organisation must locate potential service providors using either regional telephone directories or the internet. We further discuss this problem in section 3.2 and propose a possible solution.

4. PROCESSING SCENARIOS

4.1 Document Management System

The case study organisation also maintains a DMS (Document Management System) with standard processing requirements, ie. Creation and maintenance of documents which typically relate to judgment matters in a default of payment claim:

- Letter of Demand
- Issue of Statement of Claim
- Service of Statement of Claim
- Default Judgment
- Writ of Execution or Warrant
- Examination Summons
- Matter action record

Many documents may be created for a single matter, and these documents need to be retained in hard copy (and online) for a number of years (7) by law. Given this requirement, an efficient method of retrieving specific documents was identified in the case study, as well as a method of locating documents which contain details of cases with a similar pattern to a case being prepared currently. Locating documents which relate to similar matters allows the small operator to use the original documents as precedents or templates for the current case. Often, phrasing, fee structures, outsourcing agencies and methods of solution can be duplicated with very little change to the current matter. To facilitate this process, an intelligent ranking and extraction process can be used to leverage a document search engine and identify matters which relate to the search criteria.

To this end, a number of search requirements were identified in the study:

Name Search - Allows each kind of document stored in the repository to be found by its name. Long file names, spaces and foreign characters are supported by the name search method. Folders can also be retrieved using this method. In the case that the exact name is unknown, a wildcard search would be available.

Full Text Search - This is the most extensive method for conducting a document search. A user is able to search for a single word or phrase, a complex expression, or an extensive range of search standards. Misspellings

can affect the accuracy of the search, which is a major limitation, and this issue needs to be addressed.

Index Search – As the repository becomes large, this method is most suitable. Users can perform alphanumeric character searches to find index information. These searches include wild card expressions and range searches, date and date range searches to locate appropriate matters.

The mercantile agency also maintains electronic notes on matters which are filed in the document repository, the business study identified a requirement to search these notes for pattern matching criteria in order to compare outcomes and legal precedents for dealing with current matters

4.2 Location Of Costs and Fees using the Web

As a means of growing the business, the mercantile agency is looking at accepting the task of debt collection in other localities. In order to process and recover a debt, service fees applicable to the matter may relate to:

- Sheriff costs
- Process Server costs
- Solicitor costs
- Land Titles search fees
- Motor Vehicle records search fees
- ASIC (Australian Securities and Investments Commission) search fess
- Registrar General search fees

All these service providers have sliding fee schedules which require the mercantile agency to determine the fee structure which will apply when dealing with a matter in that particular jurisdiction. Court fees in each jurisdiction may also differ for a particular type of document to be lodged, and these other agencies such as court appointed Sheriffs or investigators who can locate and serve documents on behalf of the mercantile agent must be contacted and paid prior to acting on behalf of the mercantile agency. In addition, court lodgment fees are continually changing, so the source of the fee structure needs to be periodically reviewed to ensure that the current fees are being paid when court documents are lodged. For example, refer to these sites for examples of the disparity in content and style of the fee structure information available in various court localities

http://www.nt.gov.au/ntsc/doc/almanac/2004_almanac.pdf
http://www.lawlink.nsw.gov.au/sc/sc.nsf/pages/filingfees
http://www.lawlink.nsw.gov.au/dc.nsf/pages/dc_fees
http://www.courts.tas.gov.au/magistrate/circulars/cir2000_8.htm

Use of the Web to locate information relating to fee comparison, goods and services procurement, as well as access to electoral role information for skip tracing is also an expected area of growth for the mercantile industry. So the use of intelligent agents to mine the Web for relevant information is of great benefit to all businesses, but particularly relevant to a small business. The major logistical problem for the user of the Web to provide information (at this stage) is the lack of formal content standards as well as the lack of location directories for accessing the data in a structured manner.

As a means of improving office efficiencies, we propose a series of experiments using an ontology driven information system which gathers and builds a knowledge base relating to a specific subject content domain. We intend to build on Brin's Dual Iterative Pattern Relation Extraction (DIPRE) [2] method to construct a semantic ontology of domain patterns which is then used to build the knowledge base. Our system first uses a set of training pages to generate a conceptual map (ontology) of the physical training pages. The application will then use the ontology to search a text repository such as a document management system or the Web. The objective is to rank the results into a meaningful knowledge base of information relating to the sub ject domain.

5. RELATED WORK

Related work using extraction analysis methods to develop ontologies and information domains by 'trawling' the World Wide Web was begun in earnest at the same time as the Internet became a commercial entity. Research to identify phrase patterns which surround the target domain was begun by Riloff et. al. in 1996[3]. Riloff separated initial content retrieval into relevant and non-relevant domains using a frequency algorithm to determine relevance. Content deemed relevant was then used to generate patterns for subsequent retrieval runs which better filtered the content relating to the target domain. In 1998, Sergey Brin developed a more formal, iterative method for the discovery and containment of patterns (DIPRE - Dual Iterative Pattern Relation Extraction). Brin's approach has been extended by a number of research efforts including Yangarber et al. [4], Soderland and Etzioni (University of Washington) [5] and Agichtein and Gravano (Columbia University) [6] to generate domain ontologies and Q/A query facilities.

Etzioni's group generate ontology repositories using the 'Knowitall' tool they developed. Agichtein's research developed 'Snowball' which also uses an iterative approach using domain seeds to begin the extraction process. Our work is also similar to Crescenzi's Roadrunner [7] project in terms of the use of example pages to build a pattern ontology as a parser for the extracted corpus. Using this method, discovery starts with a few training pages which are known to be good indicators of the topic of interest. These patterns are used to retrieve some relevant content; the content is scanned for associated phrase patterns, and the process

is repeated, refining the target to develop the ontology or specific answer set. In our experiments, we attempt to deliver a tangible efficiency benefit to the user by applying these techniques to the business problem. We use the Web corpus to locate and acquire information on court related fees applicable to the recouping of outstanding debts in court jurisdictions within regional Australia. Our objectives in this paper are to use these techniques to develop semi automated intelligent processes which build patterns from each extraction iteration, then compare these patterns against semantic rules generated from the 'bootstrap' template.

6. INFORMATION RETRIEVAL

6.1 Prelude To Experiments

This section details the architectural design and setup for our experiments. We begin with the development of a browser based request/response interface which interacts with the user, provides input to the Web search engine, Web Service, or application specific search engine. This means the module constructs both a standard html request as well as a SOAP request. An application specific search engine will be expected to accept XML requests, the app server will deliver document URL's to the receiving interface.

When the user begins a session, initialization information is requested, such as an iteration count which limits our run, one or more learning pages which are used to create the ontology and a variable number of key words or phrases which we use to bootstrap our initial search criteria. For our purposes, we define an ontology as a conceptual image of the training pages which we store as a tree of nodes centered around the bootstrap keywords and phrases. We build our tree by first accepting and storing the users search arguments and any example pages used to generate the ontology of patterns including date and time and users id. The contents of the first extraction run are cleaned, to provide input to the pattern maker. The cleaner counts the number of hits of each keyword and passes this to the pattern maker which uses this information as a probability of usefulness. The pattern maker accepts the resulset strings from the receiving interface and clusters [8] them according to occurrence (refer to Figure 2). Patterns are created using the DIPRE algorithm for each pattern identified (keyword or phrase). Each run generates new patterns which are analyzed and compared to the ontology for relevance.

Our initial business goal is to locate and rank pages according to the closeness of the match with the training pages. The system orders the users desired information and builds a knowledge base on URL, location, fee type and fee value.

So on the first run, the discriminators (patterns) which surround the keywords and key phrases are saved and used in the second and subsequent runs. We calculate a probability for the relevance of each page returned and first try to select most obvious discriminators, then loosen constraints and select NEAR discriminators.

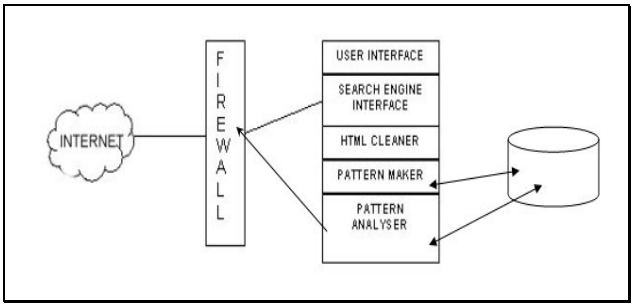

Figure 2.

7. FUTURE WORK

We report the details of our design and results of our experiments in later papers in the series. Our longer term objective is to investigate the introduction of intelligence into Web based applications as a way of improving human/computer interaction efficiencies. We focus our research in the area of small to medium business enterprises (SME's) as this market (SME's) account for more than 90% of Australian businesses [1].

Our work provides us the opportunity to:

- Research and develop applications which utilize the 'semantic' web [9] and related content presentation standards being developed.
- Research the development of agent based software using a genuine commercial environment from which to draw requirements.
- Develop leading edge software solutions which deliver tangible cost saving benefits for small businesses.

REFERENCES

[1] Collins A.J. et.al.; "Using Portal Technology for Collaborative Ecommerce"; *7th International Conference on Electronic Commerce Research*, Dallas 2004.

[2] Brin S., "Extracting Patterns and Relations from the World Wide Web", *Proceedings of the 1998 International Workshop on the Web and Databases*, 1998

[3] Riloff E. "Automatically generating extraction patterns from untagged text". *Proceedings of the Thirteenth National Conference on Artificial Intelligence*, pages 1044-1049, 1996.

[4] Yangarber R. and Grishman R.; "Extraction Pattern Discovery through Corpus Analysis";

Proceedings of the 2nd International Conference on Language Resources and Evaluation; 2000

[5] http://www.cs.washington.edu/research/knowitall/

[6] Agichtein E., Gravano L. ;"Snowball: Extracting Relations from Large Plain-Text Collections"; *Proceedings of the Fifth ACM International Conference on Digital Libraries*; 2000

[7] Crescenzi V. et. al. "The Roadrunner Project: Towards Automatic Extraction of Web Data"; *Universit`a della Basilicata – Italy*; 2001

[8] Bobrovniko D.; "Experiments in Semantic Bootstrapping with a Cluster-Based Extension to DIPRE"; *Stanford University*; 2000

[9] Berners-Lee T. et.al.; The Semantic Web; *Scientific American*; May 2001