7-25-2013

CREATE: Concept Representation and Extraction from Heterogeneous Evidence

Archana Bhattarai

CREATE: CONCEPT REPRESENTATION AND EXTRACTION FROM HETEROGENEOUS EVIDENCE

by

Archana Bhattarai

A Dissertation

Submitted in Partial Fulfillment of the

Requirements for the Degree of

Doctor of Philosophy

Major: Computer Science

The University of Memphis

August 2013

DEDICATION

In memory of my late father, Baburam Bhattarai.

ACKNOWLEDGMENTS

I am sincerely grateful to the support that have been provided to me by my advisor, Dr. Vasile Rus, along with members of my doctoral committee; Dr. Dipankar Dasgupta, Dr. King-Ip Lin and Dr. Lan Wang. I want to thank all the professors and staff at the University Of Memphis Computer Science Department for their support. I would also like to thank my friend and colleague, Nobal Niraula for always being ready to discuss on ideas and thoughts.

Finally, I would like to thank my mom, my husband, my lovely son and daughter for being the reason for my research based education, and having patience and providing support along the way.

ABSTRACT

Bhattarai, Archana Ph.D. The University of Memphis. August, 2013. CREATE: Concept Representation and Extraction From Heterogeneous Evidence. Major Professor: Vasile Rus, Ph.D.

Traditional information retrieval methodology is guided by document retrieval paradigm, where relevant documents are returned in response to user queries. This paradigm faces serious drawback if the desired result is not explicitly present in a single document. The problem becomes more obvious when a user tries to obtain complete information about a real world entity, such as person, company, location etc. In such cases, various facts about the target entity or concept need to be gathered from multiple document sources. In this work, we present a method to extract information about a target entity based on the concept retrieval paradigm that focuses on extracting and blending information related to a concept from multiple sources if necessary. The paradigm is built around a generic notion of concept which is defined as any item that can be thought of as a topic of interest. Concepts may correspond to any real world entity such as restaurant, person, city, organization, etc, or any abstract item such as news topic, event, theory, etc.

Web is a heterogeneous collection of data in different forms such as facts, news, opinions etc. We propose different models for different forms of data, all of which work towards the same goal of concept centric retrieval. We motivate our work based on studies about current trends and demands for information seeking. The framework helps in understanding the intent of content, i.e. opinion versus fact. Our work has been conducted on free text data in English. Nevertheless, our framework can be easily transferred to other languages.

TABLE OF CONTENTS

Contents **Pages**

LIST OF TABLES

Tables **Pages**

LIST OF FIGURES

LIST OF ABBREVIATIONS

CRF	Conditional Random Fields
HMM	Hidden Markov Model
HTML	HyperText Markup Language
IE	Information Extraction
IR	Information Retrieval
IDF	Inverse Document Frequency
ITS	Intelligent Tutoring System
LSA	Latent Semantic Analysis
LDA	Latent Dirichlet Allocation
MSR	Microsoft Research (Paraphrase) Corpus
NLP	Natural Language Processing
NPMI	Normalized Pointwise Mutual Information
PLSA	Probabilistic Latent Semantic Indexing
PMI	Pointwise Mutual Information
POS	Part-of-speech
RTE	Recognizing Textual Entailnment
QA	Question Answering
SVM	Support Vector Machine
T2T	Text-to-text
TF	Term Frequency
W2W	Word-to-word
WN	WordNet
UGC	User Generated Content
XML	Extensible Markup Language

Chapter 1

Introduction

"Necessity is the mother of invention"– Anonymous

1.1 New Demands in Information Seeking

As Internet has adapated itself to become the most popular information provider, there has been an exponential growth of content types, users, and applications. One of the major challenges of the 21st century is to suitably and efficiently manage this rich data and make the best use of it in various applications. Users looking for Information thus have to cope with this staggering amount of, often un-organized, digital data. This often leads them to some form of search engines as a starting point to access relevant documents with respect to input query. With this approach, the fundamental supposition is that, a user, driven by an information need, constructs a query in some query language. The query is then submitted to a system that selects those documents that match the query from a collection of documents(corpus), as indicated by certain matching algorithms. If the desired information is listed in one of the highest ranked documents, it is available within few clicks. However, in most of the cases, user has to browse through several pages and aggregate information manually. This still does not guarantee the completeness of the search result. For example, if a user wants to know the names of all countries in the world, some highly relevant documents may have a list of countries. This however, does not tell that the list is complete. Leading search engines such as Google, Yahoo, and Bing, are all based on this document retrieval paradigm. This method has proved very successful and accurate as a lot of interesting things are done to make this matching process very relevant. This approach essentially assumes that the information required is available on the web in static form. This paradigm faces a serious drawback if the desired result is not explicitly stated in a single document. The user thus has to manually browse through multiple documents and obtain the desired information. This is a tedious and often frustrating user experience. Statistics, based on previous researches, that help us understand user's search experience is summarized below.

- Around 16.2% of all page-views start from a search engine (Kumar & Tomkins, 2009).
- The average page-view for a search session has been found to be 6 in some research works (Kambhatla, 2004a). This indicates that in average, a user has to browse 6 documents before the information need is fulfilled.
- 44% of non-navigational search sessions last longer than 1 week (Kambhatla, 2004b). This indicates that a large percent of users spend time in search engines to fulfill the knowledge requirement.

The above statistics indicate a drawback of current search engines and demand a system that can essentially minimize the session time of a user. This also indicates that, in spite of having huge amount of information spread over multiple sites, users find it difficult to get desired information. Some examples of such queries could be *"countries in Asia"* or *"HP laptop reviews"*.

Another major drawback with the document retrieval paradigm becomes evident when trying to obtain complete information for real world entity such as person, company, location etc. A significant and most popular portion of web search queries are named entity queries, which target entities that can be named such as cities or persons, e.g. President Barack Obama. We present some already known statistics to support our claim.

- In an internal study at Microsoft, it was found that at least 20-30% of queries submitted to Bing search were named entities (Guo, Xu, Cheng, & Li, 2009).
  - 71% of all queries contained named entities.

- At Yahoo, it was found that more than half of all the queries contain direct reference to a structured object (SO) (Kumar & Tomkins, 2009).
  - About 52.9% of all the queries are structured objects alone.
  - 4.7% contain structured object
  - 8.5% are topic/concepts.

All these evidences lead to the imperative requirement for a new paradigm of concept retrieval as opposed to document retrieval. This paradigm is built around a generic notion of a concept which is defined as any item that can be thought of as a topic of interest. Concepts may correspond to any real world entity such as restaurant, person, city, organization, etc, or any abstract item such as news topic, event, theory, etc. It should focus on extracting and blending information related to a concept from multiple sources if necessary.

To better understand the scenario, let us start from a query which is entered by a user. The first step is to understand intent of the user. The only way to understand intent of a user, in case we do not have any background knowledge of him or her, is to understand intent of the query. Before we jump to methods to understand the intent, we will go through some fundamental query categories to better understand the demand. Broder et al., categorized queries into three types based on their intent (Broder, 2002). This category is still valid in recent research works (Jansen, Booth, & Spink, 2008). These query types are follows:

- Navigational queries: the intent is to reach a particular site.
- Informational queries: the intent is to acquire some information that may be present in one or more sites.
- Transactional queries: the intent is to perform some web mediated activity.

The targets for navigational queries and transactional queries are generally one click away from search engine results. The targets for informational queries, on the contrary, may have to face a longer session time, often making user visit numerous web pages with some or no useful information before the user finds the desired information. In the process, the user may even get fed up and quit the session without getting searched information. We further categorize Informational queries as follows:

- Factual queries: The intent is to obtain some known fact about a concept or a topic.
- Subjective queries. The intent is to obtain some form of impression of the concept or topic of other users.

The next generation of WWW will thus be driven by the attempt to understand the intent of user and trying to help user meet the intent in shortest time. Concept retrieval can answer fine-grained requirements of user queries by extracting and blending data from multiple sources. This has been predicted in a number of survey papers (Baeza-Yates & Raghavan, 2010; Dalvi et al., 2009). An approach for this would be development of a system that would process available documents and generate an assembled result specific to user query. We show some preliminary results that prove the significance of such method based on current search trend on popular search engines.

In this dissertation, we propose to develop methods to best answer informational queries, both factual and subjective. We propose separate methodologies to handle these informational queries. The details will be made more elaborate in subsequent sections.

1.2 Concept Retrieval Paradigm

Concept retrieval paradigm considers concepts and their relations as the unit of retrieval. A *concept* is defined as any item that can be thought of as a topic of interest. Concept may correspond to any real world entity such as restaurant, person, city, organization, etc, or any abstract item such as news topic, event, theory, etc. As an example, we will consider an information seeker who wants to know about Nepal. Nepal is an instance of the concept Country. The concept Country has different relations with several concepts such as Capital, National-Language, Currency, Population, etc. An information seeker may want to know about a certain relationship between two concepts, more than one relationship, or maybe all of them. Another example would be a concept related to an event *"NBA basketball game between Hawks and Lakers in Atlanta"*. The concept Event itself is abstract, but has a relationship with other concepts such as Game, Team and Location. The instance of game is basketball, teams are Hawks and Lakers, and location is Atlanta in this case. Similarly, another concept Restaurant may have relations with concepts such as Food, Ambience, and Price etc.

1.3 A Walk through Case Study on Concept Retrieval

The primary idea behind concept retrieval is to provide user with specific yet comprehensive answers based on query intent. This means that the system does all the necessary steps to identify and compile various pieces of desired answer before it provides assembled information about topic of interest. Let us walk through few examples to better understand requirements of such a system. The scope of our study is limited only to the cases where user starts information seeking process from concept retrieval based search engine. Let us consider, user enters the following query in a search engine.

Can a person lose hair due to liver problem? : fact[query1]

The text after colon, "fact" gives the system additional information that user

Table 1.1: Query response for a subjective query

Food	Positive (222)	Fresh cupcake (48), sweet icing(12), right amount(11)
Food	Negative (41)	Dry cupcake(11), cold glass(7), stale tasting(6)
Service	Positive (132)	workers nice(10), new people (9)
Service	Negative (23)	big disappointment(9), pretentious waiter(5)
Miscellaneous	Positive (129)	Good place(25), thick layer(12), near square(10)
Miscellaneous	Negative (31)	long lines(19), extraordinary demand(4)

wants factual information based on the query. This query specifically wants a yes/no answer preferably supported by some reasons. Now, the ideal task of the system is to construct a document that answers in the following way.

Yes, a person can lose hair due to liver problem.

It happens because of such and such reasons.

In another scenario, let us consider a user enters a subjective query as follows.

restaurant tasty food: review[query2]

Now in this case, the user wants to know general opinion on the restaurant named *"tasty food"* and there may be thousands of reviews for this restaurant which may be virtually impossible to read in a fixed time period. The ideal solution in such a scenario would be to present a statistical estimation of positive, negative or both the reviews in different aspects of the restaurant.

The result should be something similar to the figure shown below.

In yet another scenario, let us consider a user enters the following query

Hosni Mubarak: news [query3]

In the above query, user most probably wants to know current news on activities of Hosni Mubarak. The ideal answer for such a query for date 2/11/2011 would be as follows.

Hosni Mubarak resigned

Similarly, a user can also enter the following query *Hosni Mubarak: opinion.. [query4]*

In this case, the user preferably wants to know what general people have to say about Hosni Mubarak in current time. The ideal answer would be similar to table 1.2.

Table 1.2: Sample response to a subjective query

Resignation	Positive (222)	Obama pressures(16)
Resignation	Negative (41)	Mubarak fears (34)

The creation of such dynamic document (response to queries) covers the involvement of a wide range of technologies. We list some of the important tasks that need to be executed to develop such a complete system.

1. Considering that user starts information seeking session from a search engine, first task is to decide whether query is looking for factual or subjective information. In case of factual information, the system then needs to understand if user wants current or hot news or complete information. In case of subjective query too, the system needs to understand if topic of interest relates to a ratable product or something that is talked about in other web 2.0 technologies such as blogs, social networks etc. We broadly classify informational queries to two classes as follows.

    - Factual queries
        - Specifications
        - Recent news
    - Subjective queries
        - Real word ratable entities such as movies or products etc
        - Real world news/topic/event

With these categories of queries, user may also wish to retrieve information in a particular category.

2. The second task involves collecting documents and classifying them as opinion, rich or factual. This might be done manually or by developing an automated system that will classify a document as opinion rich or fact based using some algorithmic machinery. To further simplify the problem, we will make few valid assumptions in our work. We will consider that current news articles contain factual data whereas, comments on these articles and blog data contain opinion rich document. For a factual specification, general purpose encyclopedia such as Wikipedia is considered as it mostly well dependable facts.

3. The third task is processing all relevant documents based on query and generating result that specifically and completely answers user query. Based on the type of information required, factual or subjective, the processing of data will differ as the type of information required will be different in these two cases. For a subjective query, possible information wanted by a user will most likely be the opinion on general impression of the concept, opinion on a certain property of the concept, comparison of the concept with some other concept or a statistical estimation of user voting on certain or all properties of the concept. For a factual query, the possible information wanted will possibly be on a certain fact of concept or a list of facts meeting conditions in the query.

4. Another significant task of a good search system is information presentation to users. It should be such that user gets all needed information with minimal clicks. In other words, ideal search system minimizes distance between user query and required information. Besides, studies have shown that more interaction is bad from user experience point of view. In the current work, we

will focus more on result processing and generation rather than the presentation. Hence information visualization and presentation will be more or less out of scope for the dissertation.

The tasks (1), (2) and (4) are well studied and are hot area of current research. Task (3) has also gotten a lot of attention in the recent years. This task however, still has some important research questions to be answered. Considerable research effort is required to combine all these tasks and bring about a working system.

1.4 Web Content and its Heterogeneity

Web is the largest repository of data generated by human. The increase in the content, users and applications in web has shown exponential nature. We present some statistics to show this growth below.

- As of March, 2013 Google has indexed at least 49 billion pages [1].
- By the year 2012, 34.3% of world population started using internet[2].
- As of June 30, 2012, user growth rate is 566.4%[3].
- As of June, 2013, at least 672,985,183 web-pages exist[4].
- Around 900 million computers are directly connected to the Internet[5].
- Internet content is overall 20% redundant (Yates et al., 2007).

Web mainly contains terabytes of natural language text in several languages. It also consists of a considerable amount of multimedia files such as images and videos etc. Web pages are generally connected to other web pages with anchor texts

[1]worldwidewebsize.com, 2013

[2]internetworldstats.com, 2013

[3]internetworldstats.com, 2013

[4]Netcraft.com, 2013

[5]http://www.isc.org/solutions/survey

and links. This link structure makes web, a connected network of web pages. This link structure carries important semantic information and is advantageous for useful applications. Many researches have focused on exploiting this structure to improve or develop applications (Broder & Mitzenmacher, 2004; Leskovec, 2008). In our work, we limit our scope to text processing. It can undoubtedly be beneficial to combine the best of both to get a better system.

Based on the fact and belief aspect of content, data on the web can be coarsely categorized as follows.

- Factual data: The data that states a universal truth on a subject.
- Subjective data: The data that states personal opinion on a subject.

Factual data generally comes in the form of news articles, product specifications, online encyclopedia etc. These sources are relatively more authoritative and reliable. Due to the nature of the data, it is relatively static in nature. Factual data need not be redundant, but has to be significantly authoritative compared to opinionated data. This type of data is generated relatively in less volume compared to opinion rich data.

Subjective data, on the other hand comes in the form of user generated content(UGC) such as product reviews, social networks, blog articles and comments, etc. These sources are explicitly designed to capture user opinions in various aspects. The average quality of UGC is not as good as editorial content, however, the value of this data lies in providing a multifold opinion and varied view of the same concept and serves for different information need of users. Data is often polarized as positive and negative on the subject or more specifically on certain properties of the subject. These data are redundant and redundancy is important in such data to grasp the overall impression of the subject. Web 2.0 technology has helped increase the overwhelming amount of UGC in the web. Web 2.0 is associated with Web applications that facilitate interactive information sharing and collaboration, such as blogs, wikis, mashups and social networks. As mentioned in Ramakrishnan et al., following statistics are evident for factual and subjective data (Ramakrishnan & Tomkins, 2007).

- The amount of professional data (mostly factual) produced every day is estimated to be 2 Gb.
- The total volume of user generated content (UGC; mostly subjective) generated per day is around 10 Gb.

Factual and opinionated data serve different purposes of users information need. Due to the nature and dynamics of data, the two categories of data cannot be handled in the same way. We discuss in detail on the processing of factual and subjective data in the following sections.

1.4.1 Factual Data Processing

The first and foremost task in processing of factual data is to extract concept centric information; more specifically identification of concepts and discovery of the relations between concepts. Traditional information extraction techniques do not work well in such a scenario since we are talking about every possible concept and every possible relation between concepts. The first challenge is thus to develop an algorithm that can extract concepts and their relations. After extraction, the reconciliation of concepts is another big research area. When we talk about concepts and relationships, the idea goes on the direction of semantic web principle. We try to focus on directly applicable methods and try to make the system flexible so that it is empirical. Organizing the concept-centric data with minimal semantic information loss is yet another challenge and research area. We will talk about these methods more elaborately in subsequent chapters.

1.4.2 Subjective Data Processing

Subjective data is characterized by a personal belief on a concept or a certain aspect of a concept. This data is also expected to be highly redundant compared to factual data. Data representation and extraction in subjective data needs to be done in a different approach due to the nature and purpose of data. The first task in this case too is to identify concept upon which opinion is being advocated or opposed. Another task is to identify attributes of the concept to define or analyze the concept in detail. Other aspects such as concept disambiguation or concept-reconciliation come into play in this case too. However, when it comes to presenting response to query, the idea is slightly different in this case. In case of opinions, users are most probably trying to obtain an impression on a certain aspect of the concept of interest or whole concept or comparison of different concepts. Moreover, users may also be interested in the positive opinions or negative opinions or both.

1.5 Research Challenges Addressed

An extremely large amount of unstructured data available in web carries information that may be vital to humans. The value of this information can be reaped if there exists a translation process to modify it to some structured form. This transformation of unstructured data onto structured information requires solving many research challenges. We will briefly mention few major challenges related to the major goal of the dissertation and that have been discussed in subsequent chapters.

1.5.1 Classification of Factual and Opinion Rich Documents

As explained in the earlier section, web consists of heterogeneous data. Even when we only consider textual data of a specific language, Web 2.0 confronts us with fact based and belief based data. We handle these two types of data differently as its nature and growth follow different pattern. To identify whether the data is factual or subjective is another notable challenge. We will simplify this step making few valid assumptions. For our research work, we will only work with data sources that are explicitly fact based on belief based. For factual data processing, we will use documents in Wikipedia as the information in Wikipedia can be validly assumed to be fact based. We also experiment on medical informatics documents by crawling web-pages that are authoritative in providing medical information. The websites crawled are listed in later chapters. Similarly, for opinion, we use product and service reviews and few noted blogs that are explicitly known to be rich in opinions.

1.5.2 Generic Concept and Relation Extraction

A structured representation of textual information can be directly beneficial for knowledge discovery. For the factual data, identifying a word or a phrase that refer to a real world concept is the foremost challenge that needs to be solved at data extraction and storage time. Since we are talking about every possible concept and all of its relations, a constrained domain dependent method becomes limited. This indicates the need of a method that can generically identify any unseen concepts and relations based on some heuristics. We propose and show two major ways for generic concept and relation extraction.

1.5.3 Subsumption Hierarchies

Concepts naturally form a hierarchy. This phenomenon is about the methods that help bring about hierarchical association in concepts. Starting with a concept itself, it is a superset of all the items that we deal with. Taking the geographic view of the world as an example, it can be divided to continents, the continents to countries, countries to cities/villages, etc. Organizing concepts in such hierarchical form can help in knowledge representation. This form also makes the process of inference relatively straightforward in some aspects.

1.5.4 Query Processing

Query is one of the most reliable gateways to predict the intent of the user. Thus it is very important and subtle to understand it before any processing is done.

It is also a well studied research area and consists of significant work in its own. In this work, we will only focus on the concept aspect of the query and will ignore the intent understanding part in the dissertation.

1.5.5 Scalability

We are attempting to exploit large text corpora for extracting the desired generic knowledge. Moreover, the methods we described above are computationally intensive. Bringing a balance in efficiency and accuracy becomes another research challenge. This would require algorithms that can handle scalability with techniques of parallelization, distribution and dynamicity. We propose to use systems that have considered these factors: Lucene, MongoDB to name a few.

1.6 Our Approach

The fundamental problem we propose to solve in this dissertation is the dynamic generation of a response for user queries. This gears towards the huge goal of paradigm shift of concept based retrieval. We thus describe our system as **"a generic concept based unsupervised system"**. Our system bases on the idea of minimal requirement of human supervision. The system handles factual and subjective data differently. We build a system over an unstructured generic text. For the factual data, we first transform text to structured concepts and relations, derive the hierarchy and use inference to make the system more resilient and accurate. The system starts from few hand coded rules and then learns from the data it obtains. With this design, the system can make itself better overtime, and in the long run, the system will become self-contained. Similarly, for subjective data, system identifies significant aspects/properties of the concept. Based on these aspects, the system then generates a statistical estimation of wisdom of crowds on target concept with supporting textual cues. The major properties of combined system would thus be significantly automated, domain independent and efficient. We briefly describe the overall framework in Figure 1.1. below.

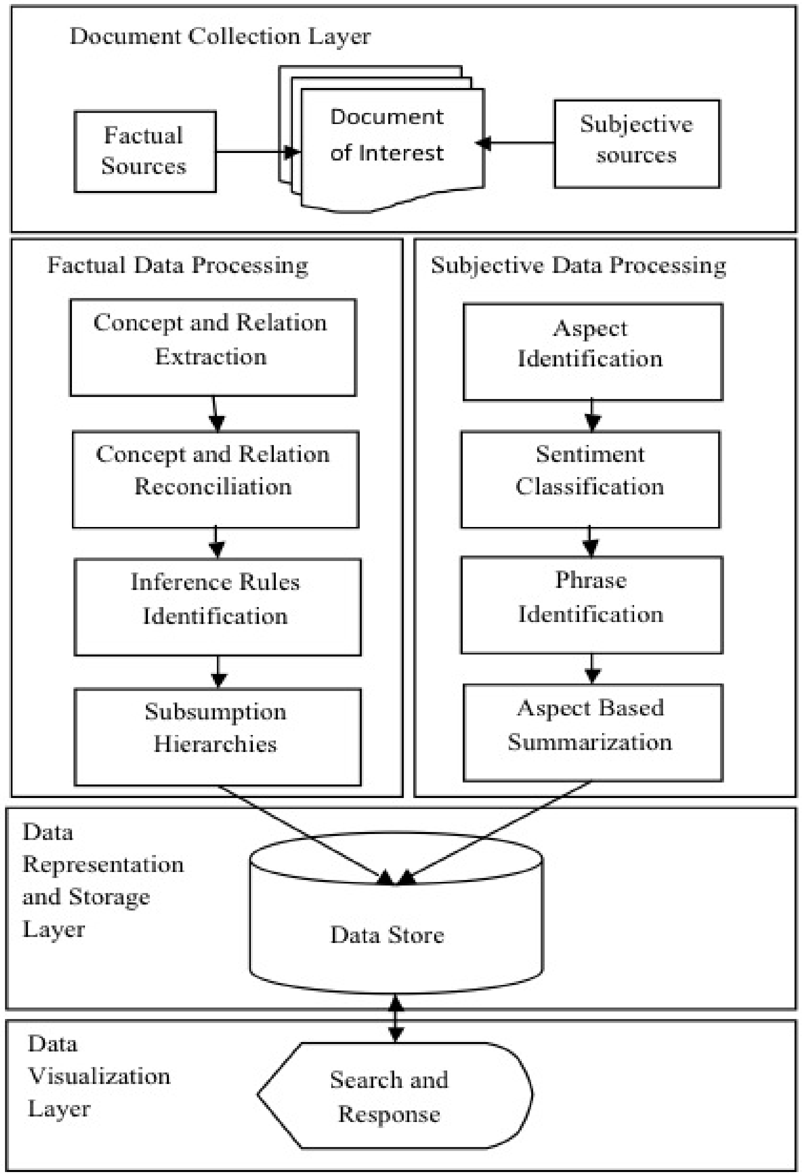

Fig. 1.1: Overall System Architecture

1.6.1 State of the Art

Our vision is to build an automated system that can read textual data to a deeper extent compared to bag of words model. In this sense our work is similar to the works of Carlson and his group (Carlson et al., 2010), Dalvi and his group (Dalvi et al., 2009) and Textrunner (Yates et al., 2007) etc. Most of the systems, however, work towards factual text reading. Carlson et al start with a handful of examples of each predicate in the ontology (Carlson et al., 2010). They use semi-supervised bootstrapping approach to continuously read and update knowledge base with an Expectation Maximization like algorithm (Carlson et al., 2010). This system, however, is limited to the concept list and hence cannot cover every possible concept and relation available in web. Dalvi et al, advocate for the requirement of new paradigm of web of concepts (Dalvi et al., 2009). They support their belief with statistics that demands this new paradigm. They also outline different fields that could benefit from this paradigm and also note the challenges that need to be solved to visualize this system. Yates and his colleagues also work towards the same goal by introducing TextRunner, a system capable to extract entities and relations at web scale (Yates et al., 2007). Baeza-Yates et al foresee newer generation of web search which is based on concept/entity retrieval (Baeza-Yates & Raghavan, 2010). Their theme of newer search system is guided by the principle that people search not to search, but to complete some task. Hence they infer that the search principle should help users complete their tasks as efficiently as possible. The linked data consortium, a brainchild of the inventor of the World Wide Web, Tim Berners Lee, is another early initiative to move the web community towards this paradigm. We would differentiate our approach with the linked data methodology in method of generating structured data for the purpose. The linked data method supposes that human expertise can be used to meticulously define the ontology. Our approach, on the other hand, starts from the unstructured text that is already available in the web and works through the development of this automated system. Other systems focus on more structured part of large factual collections such as Wikipedia (Auer et al., 2007; Suchanek, Kasneci, & Weikum, 2007; Weld, Hoffmann, & Wu, 2009).

1.6.2 Objective of the Dissertation

The dissertation advocates for the huge paradigm shift towards concept based information retrieval. We propose an approach where different genre of text is processed differently in contrast to traditional document based information retrieval. With this broad objective in mind, we will define few specific objectives here.

1. Development of a framework for concept based fact retrieval.
2. Development of a framework for concept based opinion retrieval.
3. Development of a working interface for overall concept based information retrieval.

Contributions from Objective 1

Development of concept based fact retrieval involves a number of research challenges. We propose to work on major fundamental challenges. Our contributions on this objective is as follows:

1. A mechanism to transform unstructured data to structured form with minimal human supervision.
   - Structured data generation based on lexico-syntactic patterns.
   - Structured data generation based on bootstrapping methods such as DIPRE (Brin, 1999) and Snowball (Agichtein & Gravano, 2000).
2. A mechanism for hierarchy development in an unsupervised way.
   - A mechanism to evaluate the confidence of accuracy for the system.

Contributions from Objective 2

Development of concept based opinion retrieval too, has to cross notable research hurdles. We propose to develop a generic framework for this purpose that can accurately extract opinion related information. Our contributions on this objective will be as follows:

1. A mechanism to extract the aspects of a concept in an unsupervised way.

- Aspect extraction based on probabilistic topic models such as LDA.

2. A mechanism to detect the sentiment polarity of a text of interest in an unsupervised way.

   - Sentiment identification based on user ratings
   - Sentiment identification based on Pointwise Mutual Information(PMI).

3. A mechanism to extract representative phrases related to each aspect of concept.

   - Phrase extraction based on lexico-syntactic patterns.

4. A mechanism to perform statistical estimation of representative opinion phrases.

Contributions from objective 3

Development of a working interface of overall concept based retrieval involves efficient design and implementation of a distributed system. Our contributions on this objective will be as follows.

- Development of a web based application that can be used for manual evaluation.
- Design of concept based indexing system based on traditional inverted index and relational database concept.
- Manual evaluation of developed concept based search engine prototype.

1.6.3 Delimitations of the Dissertation

We speak about a huge leap from document based retrieval to concept based retrieval paradigm. The scope of our work for dissertation however, does not cover every necessary task in this new paradigm. We outline some of the notable limitations of our work here.

1. We do not develop a mechanism to classify documents from the web as factual, news based, opinion rich documents. We instead make few valid assumptions for factual and opinion rich document categories.

2. We do not build a framework for a complete ontology that is capable of working with complex natural language concepts such as synonymy and polysemy.

3. We do not explore the properties of extracted relations that provide additional knowledge about relations such as symmetry, functionality, injection, transitivity, reflexivity etc.

1.6.4 Dissertation Outline

The rest of the dissertation is organized as follows. In chapter 2, we introduce and define Open Domain Information Extraction, the most prominent and fundamental problem for concept centric information management. We also present methodological and technical foundation to this task based on related work in the open domain information extraction research community. Chapter 3 introduces our generic concept and relation extraction system based on lexico-syntactic patterns. It also presents preliminary results and studies on its limitations. Chapter 4 presents the framework of iterative pattern induction based concept and relation extraction. This chapter also introduces our prototype: CREATE, which is based on pattern induction paradigm along with the experimental results of this prototype. Chapter 5, further explores on information management of extracted relational tuples. Chapter 6 presents concept based opinion retrieval model and related work and the experimental results for the system. We will discuss about applications and open questions in chapter 7. Chapter 8 summarizes our work and findings of it and points out potential areas to explore and as a future work.

Chapter 2

OPEN DOMAIN INFORMATION EXTRACTION

"Facts are many, but the truth is one". – Rabindranath Tagore

2.1 Introduction

A diverse field of researchers have been working towards the same goal of development of a system that is capable of processing available documents and generate an assembled result specific to user requirement. Research communities in database, computational linguistics, semantic web and artificial intelligence have been investigating this problem with their own views and different goals. When talking specifically about structuring the web, there are two broader categories of researches that work towards this idea. Semantic web community (Bizer, Heath, & Berners-Lee, 2009) takes the top down approach of creating and defining formal ontologies (schema) first and then populating data based on the schema. This often involves human intervention and is highly labor intensive and to completely cover web scale information is a big challenge. The big advantage is that the information is highly aunthentic and disambiguated. Information extraction community, on the other hand, takes the bottom up approach of extracting semantic unit of information from web, without any schema in hand, and then subsumes the information into lighter structural form. This process is mostly automated and has a higher risk of producing noisy and ambiguous data.

Having said that, since the overall goal in the dissertation is to use minimal human supervision, we choose to focus in open domain information extraction. We will start by defining open domain information extraction and discuss different methodologies that have been implemented for this process. We will first define information extraction in general. The process of conversion of data from unstructured to structured form is popularly known as **Information Extraction**(IE). Information extraction brings in itself, different challenges and advantages to the information management community. Moens et al define Information Extraction as follows (Moens, 2007):

Information extraction is the identification, and consequent or concurrent classification and structuring into semantic classes, of specific information found in unstructured data sources, such as natural language text, images, audio and video, providing additional aids to access and interpret the unstructured data by information systems.

Structured information representation and extraction is an active area of research. Information extraction can be seen as a two step process, first, the entity recognition task and second, the relation extraction task. Research work on information extraction started as early as 1980s (Defense Advanced Research Projects Agency Report, 1987; Defense Advanced Research Projects Agency Report, 1993). A primary problem in this task is to develop an extraction model for the extraction task. These extraction models fundamentally learn or identify extraction rules/patterns to extract entities and relations from unseen text.

Traditional information extraction methodologies tend to extract a predefined relation between named entities annotated in a different process. Existing work on pre-defined relation extraction have implemented methods of supervised, semi-supervised, bootstrapped and unsupervised classification (Zelenko, Aone, & Richardella, 2003; Kambhatla, 2004c; S. Zhao & Grishman, 2005; Bunescu & Mooney, 2006). Supervised relation extraction methods are based on the availability of labeled training data. While this method might be useful and accurate for smaller data with limited entity types and relations, it cannot scale to extract entities and their relationships in web due to the sheer volume and heterogeneity of data. For open information extraction methods, since they do not have predefined relations, it is very hard, if not impossible to generate labeled data for all potential relations in large text corpora.

2.2 Definition

Open domain information extraction is the process of extracting relational tuples of the form *(concept1, relation, concept2)* by making a single or iterative passes over the natural language corpus without requiring relation specific training data. We define relation tuple as the smallest unit of semantic information that is machine readable. This process, helps extract entities and the relation between these entities in text corpus. A general model of most of the open domain information extraction systems adhere to following three steps.

1. Label seed or training data manually or with some heuristics.

2. Develop a model of relation extractor.

3. Extract more relation tuples from unseen sentences.

Figure 2.1 portrays a general open domain tuple extractor.

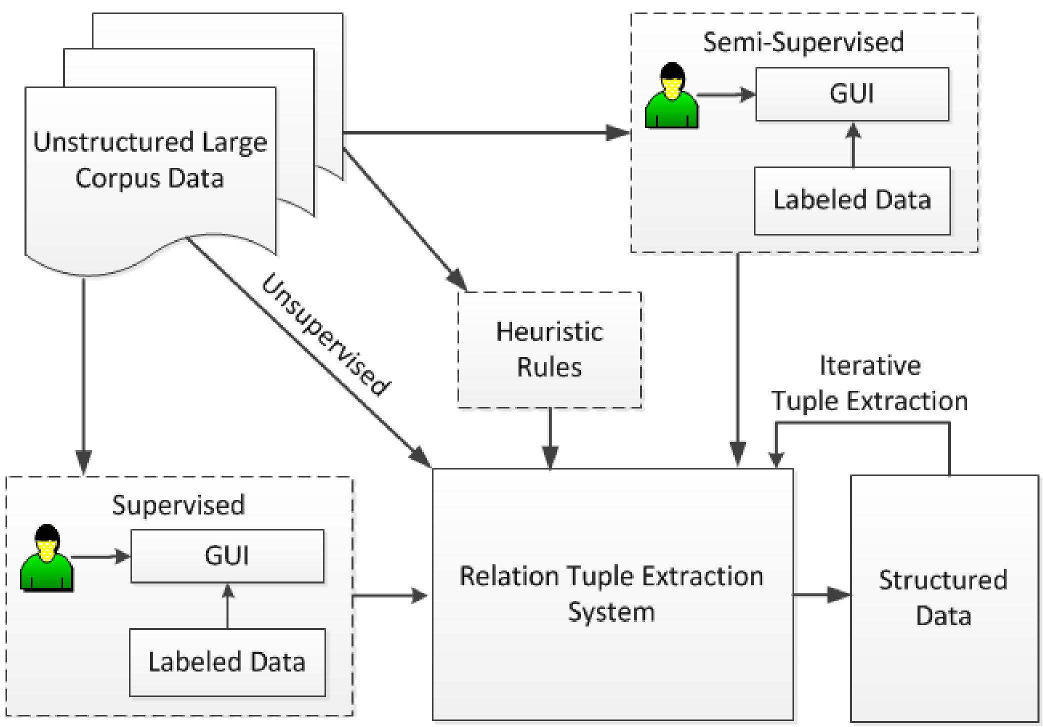

Fig. 2.1: A general view of Open Information Extraction System

2.3 Methodologies

One of the major goals of open information extraction is to build automated system that can read textual data to a deeper extent compared to bag of words model. The development of such a model traces its path from sophisticated knowledge engineering and learning based systems. Knowledge engineering approach involves constructing, testing and updating extraction patterns manually after meticulously studying a portion of representative corpus. This method can build highly accurate systems given that the rules are developed by domain experts and are fairly general. The learning based systems, on the other hand, learn the patterns from the text itself. This method, however, requires a few or considerable labelled data to learn the patterns.We will go through some prominent works in recent years categorized by their method in the following section.

2.3.1 Resource based Methods

Resource based methods are specific to a particular data source such as Wikipedia [1] or PubMed articles[2]. Most of the work done in this branch have exclusively used Wikipedia for information extraction (Auer et al., 2007; Suchanek et al., 2007; Wu & Weld, 2010). These methods focus on the structured part of Wikipedia to extract tuples based on Wikipedia-centric properties. The WOE systems (Wu & Weld, 2010) introduced by Wu and Weld make use of Wikipedia as a source of training data for their extractors, which leads to further improvements over TEXTRUNNER (Yates et al., 2007). The idea is that wikipedia articles are organized such that each article in Wikipedia is about a primary object and the articles mostly contain infoboxes which are structured summaries of the article. They propose a self-supervised classification system that learns from the relation tuples extracted from infoboxes. These systems have claimed to achieve very high

[1] http://en.wikipedia.org

[2] http://www.ncbi.nlm.nih.gov/pubmed

level of precision and recall. After developing a model based on infobox tuples, the system can then successfully learn tuples from unstructured articles in wikipedia. Yago is another similar system that bases off Wikipedia properties (Suchanek et al., 2007). It uses some manually written rules and some heuristics to extract structured information from infoboxes and other structured data such as Wikipedia category pages etc. Yago, unlike IWP(Intelligence in Wikipedia Project) limits the relations to few specific relations such as type, subclass-of, means, bornInYear, diedInYear, establishedInYear, locatedIn, writtenInYear, politicianOf and hasWonPrize. The infoboxes in Wikipedia contain a lot of useful information and attributes about the object. However, it still misses a lot of important information written in natural text. Moreover, these systems by their nature are not transferable to other domains such medical informatics or tutoring or something else. Yan et al. used the characteristics of wikipedia and performed clustering of patterns to extract relations without human supervision (Yan, Okazaki, Matsuo, Yang, & Ishizuka, 2009). They report a precision as high as 84% with deep linguistic parsing. Other works also use Wikipedia for ontology development for entities (Syed & Finin, 2010). Min et al extract relation tuples based on entity similarity graph and pattern similarity (Min, Shi, Grishman, & Lin, 2012).

2.3.2 Parsing based Methods

An intuitive approach to most of the semantic understanding of natural text at sentence level is to perform parsing of the sentence to understand the structure and meaning of the sentence. This structure obtained after parsing can be used for identifying domain independent relationships between entities. However, the problem with parsing based system is that it takes significant amount of time and resource to get a complete dependency structure. Banko and her colleagues reported that 2.5 million articles in Wikipedia (as of 2008) consisting of a total 1 billion words would require 23 CPUs to parse all of the sentences in the English Wikipedia corpus in a single day with Minipar, and over 8000 CPUs using Klein and Mannings parser (Banko, Cafarella, Soderland, Broadhead, & Etzioni, 2009). This suggests that parsing based methods are not easily web scalable. Moreover, the hand crafted rules that is used to extract relation tuples from parsed sentences may not cover many cases with which tuples could be extracted. The DIRT system introduced by Lin and Pantel, works in an unsupervised way using Minipar parser to produce relation tuples (Lin & Pantel, 2001). The major goal of DIRT is to discover synonymous relations broadly used for inference in natual language text. As an example it tries to detect *"X is author of Y"* is same as *"X wrote Y"*. The system extracted 231,000 unique tuples from 1GB collection of newspaper text. The authors of DIRT discovered that accuracy of their system was dependent on relation of interest. Another similar system Knext applies a set of human crafted general rules to parse tree of sentences (Schubert & Tong, 2003). There were around 80 rules applied to parsed sentences. On average, this system extracted 2.47 propositions and after detecting duplicate propositions, it extracted 1.78 propositions on average. It used Penn Treebank for tuples extraction as it can be considered fairly general for open domain. Similarly, Clark and colleague also used parsed structure to extract relational tuples which were extracted from the structure subject-verb-object in the dependency tree (Clark, Harrison, & Thompson, 2003).

2.3.3 Knowledge Engineered Systems

Knowledge Engineered Systems is a method that uses human expertise to engineer the rules and patterns to extract entities and relations. This method is also used as an initial seed generation step or a training data generation step in the learning based systems. The major challenge in this method is the amount of exercise required to formulate those rules. Moreover, these rules cannot be complete as natural language has myraids of ways to express an idea in the form of a sentence. It might be difficult to verify the completeness and generality with such manually defined rules. This process is made easier with interface tools such as GATE (Cunningham, 2002). Another similar framework is UIMA that provides support for preprocessing pipelines to help in knowledge engineering (Ferrucci & Lally, 2004). Other open source systems that facilitate the pre-processing step include but are not limited to libraries from the Stanford NLP group, and several others listed under the OpenNLP effort.

2.3.4 Learning based Systems

These systems try to learn the extraction patterns and rules automatically either starting from a small size of seed patterns or building a model from a pre-annotated learning corpus. Some of the early learning based Information Extraction Systems learned the extraction rules from the training data (Riloff, Wiebe, & Wilson, 2003; Craven et al., 2000). These systems had a big disadvantage of requiring a sizeable amount of training data to build. This led to the idea of weakly supervised methods which only required a handful of seed corpus to start with and they continually learned and expanded their patterns to extract more entities and relations (Brin, 1999; Agichtein & Gravano, 2000).

Another line of work that based on developing unsupervised systems did not even require a seed set to start with (Lin, 2003; Ciaramita, Gangemi, Ratsch, Šarić, & Rojas, 2008; Bunescu & Mooney, 2007). These methods exploited the information exhibited by dependency and constituency parsers to identify concepts and corresponding relational words/phrases. We will discuss about each learning based methods in the following sub-sections.

2.3.5 Semisupervised Learning based Methods

Semisupervised methods start with a few manually provided domain independent extraction patterns that will extract training tuples. These training tuples are generally extracted based on the dependency parsing of the corpus. Using these training samples, sequence based classifiers are trained and more tuples are extracted. The first Open IE system was TEXTRUNNER, which used a Naive Bayes model with unlexicalized POS and NP-chunk features, trained using examples heuristically generated from the Penn Treebank (Yates et al., 2007). Subsequent work showed that utilizing a linear-chain CRF (Banko, Etzioni, & Center, 2008) or Markov Logic Network (Zhu, Nie, Liu, Zhang, & Wen, 2009) can lead to improved extraction.

2.3.6 Unsupervised Learning based Methods

Unsupervised learning based methods logically do not require any kind of human input and evaluation for the system to perform. It tries to extract relational tuples based on methods such as clustering, parsing etc. With an effort to develop an unsupervised system, Shinyama and Sekine proposed a technique which they named unrestricted relation discovery which discovers all possible relations from natural text, and then presents them as tables (Shinyama & Sekine, 2006). The idea employed was to cluster articles that contained entities bearing similar relation to each other such that parameters of same relation in the same role remain in one column. In each cluster, the system then performs named-entity recognition, reference resolution and linguistic parsing which is used to generate relational patterns. Due to the pairwise vector-space clustering approach, this system is not scalable to all possible relations in web size data.

2.3.7 Self-supervised Methods

The first information extraction system that attempted to implement a general approach that targeted all entities and relations was introduced by Yates and his colleague (Yates et al., 2007). Their system, called TextRunner, is an open information extraction system as it can extract almost all entities and their relations. They extract relational tuples of the form (r, e1,., en), where e1,.en indicates entities and r represents a relation among the entities. For example, from the sentence “Microsoft is headquartered in beautiful Redmond” (from (Banko et al., 2008), the tuple extracted is (is headquartered in, Microsoft, Redmond). TextRunner (Yates et al., 2007; Banko et al., 2008) uses Nave Bayes and then Conditional Random fields (CRF) to identify relations and entities. Since CRF is a supervised machine learning algorithm, they developed a method to create training data using relation independent rules. Rules they developed were based on syntactic parsing; specifically, phrase parsing methods. They determined positive and negative examples based on these rules and then used the examples as training data for CRF. Before using CRF, the method first applied a phrase chunker to identify noun phrases, which were considered candidate named entities. These candidate entities should be within a given distance apart, in terms of words, in order to be considered part of a relation. The TextRunner, while being powerful than traditional information extraction methods still has several limitations. It assumes that a relation between entities always occurs in between them, which is not true in many cases. For example, in the sentence *"Google and Youtube merged"*, the relation *"merge"* does not appear in between Google and Youtube. Here, it also ignores the role of modifiers, which are words modifying other words such as adjectives, that can define or specialize a concept or relation. For example, in the sentence *"Nepal is a landlocked country"* TextRunner identifies the following tuple (isa, Nepal, country). This type of mapping from the original sentence to the tuple leads to the loss of the fact that Nepal is landlocked. An ideal system should be able to transform information in unstructured data to structured form without loss of meaning and knowledge.

2.3.8 Probabilistic Topic Based Methods

Probabilistic topic based models have also been used to infer relation between entity-pairs (Chang et al., 2009; Yao, Haghighi, Riedel, & McCallum, 2011). These models assume relation tuples as atomic observations in documents rather than word observations in standard LDA model.

Table 2.1: Comparison of Notable Open Information Extraction Systems

System	Itr	Pattern	Machine Learning	Human Labeling	Features	Text Source
Nubbi (Chang, Boyd-Graber, & Blei, 2009)	1	N/A	LDA variant	N/A	Lexicon	bag of words for entity context and entity pair context
Statsnowball(Zhu et al., 2009)	1	lexical pattern/pos pattern	Markov Logic Network	Seed Data Annotation	Lexicon, POS tag, Chunk	Any unstructured text
KnowitAll(Etzioni et al., 2004)	1	Hearst like patterns	Naive Bayes	No human involvement	Lexicon, POS tag, Chunk	Any unstructured text
TextRunner(Yates et al., 2007)	1	heuristic patterns for dependency parsed text	Naive Bayes, Conditional Random Fields(CRF)	Developing heuristic rules	Lexicon, POS tag, dependency parsing	Any unstructured text
WOE(Wu & Weld, 2010)	1	wikipedia infobox patterns	Conditional Random Fields(CRF)	N/A	dependency parse, POS tag	wikipedia infobox for training
Nell(Carlson et al., 2010)	>1	contextual patterns	L2 - regularized logistic regression	Seed Data Annotation	Lexicon, POS tag	Any unstructured text
Reverb(Fader, Soderland, & Etzioni, 2011)	1	adjacent entities and relation	Logistic Regression	needed for confidence calculation of tuple	Lexicon, POS tag, Chunk	Any unstructured text

2.4 Comparison of Notable Open Information Extraction Systems

The direct comparison of methods of open domain information extraction is difficult because of many reasons. One major reason is that extraction tuples are highly subjective and even inter-human agreement could have high variance. Additionally, evaluating tuples syntactically as well as semantically is also very difficult. We compare some of the prominent open domain extraction system in table 2.1.

2.5 Conclusion

The major goal of open domain information extraction is to extract the key facts in natural text with an automated process. Various methods ranging from manually developed rules to resource based methods, supervised, semi-supervised, unsupervised and pattern induction based methods have been applied to perform extractions. We made an attempt to define open domain information extraction in earlier sections of the chapter. We then summarized few of the notable methods in each category to make a background of methods of related work which will further help in understanding following chapters.

Chapter 3

DEPENDENCY BASED RELATION EXTRACTION

"All truths are easy to understand once they are discovered; the point is to discover them". – Galileo Galilei

3.1 Overview

The first step towards the paradigm of concept based retrieval involves generating structured or machine readable data from freely available documents. One of the preliminary requirements for factual data is to extract generic concepts, and their relations. Knowledge is assimilated in knowledge base in the form of concepts and their relations. As we see in this sentence, *"Nepal is a country"*, our goal would be to extract structured information in the form of tuple: (isa, Nepal, country). This is, however, not an easy task as it requires the system to understand the semantics of the sentence, which cumulatively constructs the overall meaning of the containing document.

In this chapter, we will propose a new algorithm which extracts generic concepts and relations, with high precision and recall. This initial step will significantly help in bridging the gap between the applications that build over structured data and unstructured data, abundantly available in the web. This also fits very well in our broader vision of concept based retrieval paradigm. This algorithm can be equally advantageous to other systems, such as automated tutoring in automated concept-map creation, automated question answering system, etc.

3.2 Problem Description

As mentioned earlier, our goal is to automatically extract concepts and their relations. To this end, we view a relation as a binary function between two concepts. In contrast to other approaches, this method does not need any annotated instances. However, this approach demands to formulate a small set of maually developed rules to extract concepts and their relations. We propose to develop the rules on top of dependency trees of input sentences. In addition, we could also use existing dictionary, e.g. Wordnet (Miller, 1995) to infer similarities between relations.

The task can be formalized as in following. We have a corpus of interest, denoted by C.

The corpus C is a bag of documents such that, each document, $d \in C$

A document d is considered to be a bag of sentences such that :

$d = s_1, s_2, ., ., ., ., s_n$

A sentence, 's' is represented as a set of tuples of the (relation, concept1, concept2) that indicates the relation between the two given concepts. A sentence S is thus a conjunction of tuples:

$S = t_1 \cap t_2 \cap t_3 \cap ... \cap t_4$ where a tuple, t_i is binary function of the form.

$t_i = rel(concept_1, concept_2)$

A query Q, is also represented in a clausal normal form.

$q = t_1 \cap t_2 \cap ... \cap t_x$

where one of the values, participating concept or relation is not known. The tuple would be something like:

$t_i = rel(?, concept_2)$, or,

$t_i = rel(concept_1, ?)$, or

$t_i =?(concept_1, concept_2)$

The task is to find all sentences that have the given query in them and return the list of sentences to user. Matching a query to existing tuples may involve some sophisticated matching operations. For example, for a query $isa(x, country) \cap propertyof(x, landlocked)$, the corresponding result would be all the values of x that matches the above tuples. As an example, matching tuples would be isa(Belarus, country) and propertyof(Belarus, landlocked). Note that the tuples need not be in the same sentence, or in the same document. This illustrates the strength of concept-based retrieval paradigm for cases in which although the facts are mentioned in different sentences, we can still infer the property based on relations of the concept.

While this particular query may have a page that specifically answers it, our method does not require it to be specified in the same document. To evaluate this, we purposely removed the list of countries page from Wikipedia articles and obtained the list of countries in the world. Our result will be evaluated in the results section.

3.3 Preprocessing: Lexical Analysis

Document preprocessing step is an important part of all natural language processing task. Our preprocessing steps involve Tokenization, Lemmatization, POS tagging etc. Tokenization involves breaking down sentences into words, which often begins with a preprocessing step that removes unimportant characters from the input text. Lemmatization involves transforming words of different morphological forms, such as in different tense, to a single form. For example, the lemmatized form of words “sold” and “sell” is “sell”. Part of speech (POS) tagging is the process of labeling each lexical token with a syntactic type. Other preprocessing steps are stopword removal where insignificant words removed from the text of interest. A word is considered stopword, if it occurs in more than 80% of all the documents of interest.

3.4 Dependency Parsing of a Sentence

A major step in the transformation of unstructured text to structured form is the generation of concept-relation tuples for each sentence. This obviously requires a deep textual analysis to determine the concepts and their relations. Since we are interested in developing a system that requires minimal human intervention, we want to use domain-independent methods to perform concept and relation extraction.

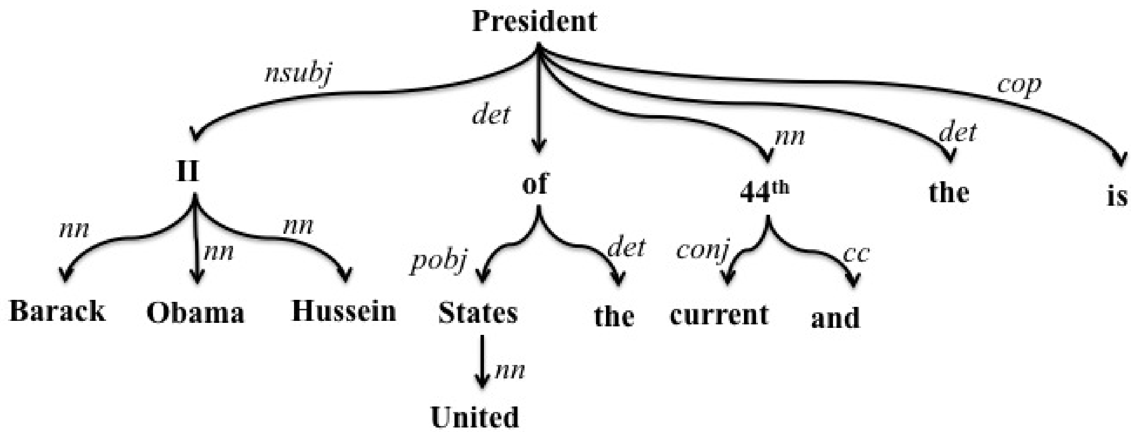

Fig. 3.1: Dependency tree of sentence *"Barack Hussein Obama II is the 44th and current President of the United States"*

One of the popular strategies to understand linguistic structure is dependency grammar parsing. This grammar is well suited to understanding the semantic relation between words as it presents the relations in a linked structure. A sentence can be transformed to a dependency tree of words using language parsers such as MINIPAR (Lin, 2003; De Marneffe, MacCartney, & Manning, 2006) etc. With such a representation, the grammatical relationship of words can be easily understood. The method represents all sentence relationships as a set of dependency relations in triplets where a triple expresses a relation between a pair of words. In our preliminary work, we used the Stanford dependency parser since the relations between words are relatively complete and yet the parser is fast (De Marneffe & Manning, 2008). The Stanford Dependencies(SD) are triplets of the following form: (name of the relation, governor, dependent). For example, a dependency tree of the sentence "Barack Hussein Obama II is the 44th and current President of the United States" is shown in figure 3.1. The dependency relations for the sentence in figure 3.1 is given by the Stanford parser are given in Table 3.1.

We propose to utilize the relations given by this dependency tree to generate concept-relation tuples based on domain-independent patterns developed in our process. We will explain the patterns and the method in detail in later sections.

Table 3.1: Dependencies of the sentence in dependency tree in Figure 3.1

Relation	**Governor**	**Dependant**
nn	ii-4	barack-1
nn	ii-4	hussein-2
nn	ii-4	obama-3
nsubj	president-10	ii-4
cop	president-10	is-5
det	president-10	the-6
nn	president-10	44th-7
cc	44th-7	and-8
conj	44th-7	current-9
prep	president-10	of-11
det	states-14	the-12
nn	states-14	united-13
pobj	of-11	states-14

3.5 Lexico-Syntatic Patterns for Concepts and Relations Extraction

As discussed earlier, dependency parsing helps detect syntactic relations between words in the form of dependency relations. As a proof of concept, we used dependency information onto general lexico-syntactic patterns that allowed us to extract concepts and their relations from a given input sentence. The extracted relations were then represented as binary functions with two concepts as arguments.

For the sentence mentioned in the above section, the relations that are to be extracted from the dependency tree are:

$isa(BarackHusseinObamaII, 44thPresident)$

$isa(BarackHusseinObamaII, currentPresident)$

$44thpresidentof(BarackHusseinObamaII, UnitedStates)$

$currentpresidentof(BarackHusseinObamaII, UnitedStates)$
We carefully studied the dependency structure of few sentences and generated rules to extract domain independent concepts and relations from such dependency structures. To make the system complete, we studied each dependency relation as given by the Stanford parser (De Marneffe & Manning, 2008) and used the ones that seemed to be useful to extract concepts and their relationships. The algorithm we developed to extract the tuples is given in Table 3.7.

After we extract the tuples from the sentence, we represent the sentence as a conjunction of concept-relation tuples. With such a representation, much more flexibility is added in the system as the system no longer has to rely on keyword importance and document retrieval. It can now deal with concept representation, extract all relevant relations to a concept of interest and thus create a dynamic response to query. Transforming unstructured text to structured form, however, does not solve the problem completely. Natural language has inherent features such as synonymy, polysemy, common sense inference etc. This will be a part of our long term goal where we will use structured data as a basis.

An important preprocessing step before trying to extract relations between concepts is to identify concepts that may be present in the form of a compound noun or collocation. Concepts may appear as composed of multiple words (e.g. San Francisco) or single words (e.g. California). For example, in the sentence: *San Francisco is the fourth most populous city in California*, the words San and Francisco have a combined meaning which is indicated by dependency relations nn produced by the Stanford parser. In the preprocessing step, we create a new word composed of individual words glued with an underscore. The preprocessing algorithm presented in Table 3.2 performs the preprocessing of noun phrases.

Conjunction words exhibit a special property. If two words are connected

Table 3.2: Algorithm for preprocessing noun phrases in sentence

Preprocess Noun Compound and Numeric Modifier
$\forall\, rel(c_1, c_2)\, where\, rel = nn \cup num \cup number,$
$\quad nounPhrase = c_2 + c_1, concatenating\, c_2\, and\, c_1$, in ascending order of number associated with words.
$\forall rel(c_3, c_4) where\, rel \neq nn \cup num \cup number$,
$\quad if(c_3 == c_1 \| c_3 == c_2),\ replace\, c_3\, by\, nounPhrase$
$\quad if(c_4 == c_1 \| c_4 == c_2),\ replace\, c_4\, by\, nounPhrase$

Table 3.3: Algorithm for preprocessing of *"and"* conjunction in sentence

Preprocess Conjunction
$\forall\, rel(c_1, c_2)\, where\, rel = conj,$
$\quad if\, \exists\, rel(c_1, c_3)\, where\, rel = subj\, \&\&\, \nexists\, rel(c_2, c_4)\, where\, rel = subj,$
$\quad$ add tuple $subj(c_2, c_3)$ to dependencies and vice-versa
$if\, \exists\, rel(c_1, c_3)\, where\, rel = obj\, \&\&\, \nexists\, rel(c_2, c_4)\, where\, rel = obj,$
$\quad$ add tuple $obj(c_2, c_3)$ to dependencies and vice-versa
$if\, \exists\, rel(c_1, c_3)\, where\, rel = prep\, \&\&\, \nexists\, rel(c_2, c_4)\, where\, rel = prep,$
$\quad$ add tuple $prep(c_2, c_3)$ to dependencies and vice-versa

with a conjunction, these words share some common properties. For example, in the sentence: *Kathmandu is the nation's capital and the country's largest metropolis.*

The concepts *"capital"* and *"metropolis"* are related with a conjunction meaning they share some common property. If the concept *"capital"* forms a tuple isa(Kathmandu, capital), then the concept *"metropolis"* will also form a tuple isa(Kathmandu, metropolis). This preprocessing step is handled in the algorithm given in Table 3.3.

After compound names and conjunction preprocessing, we will extract concept relation tuples. Concepts usually correspond to a dependent of subject or object relations.

Table 3.4: Algorithm to extract *"isA"* relation from a sentence

extract isA relation
$if(\exists\, rel(c_1, c_2)),\, where\, rel = subj$ &&
$\exists rel(c_1, c_3)$, where rel=cop, then
create a tuple t, where
$concept1 = c_2$,
$concept2 = c_1$ and
$relation = isA$

One of the special relations we discover is *isA* relation which is obtained from a copular verb. This relation most of the time represents the instantiation of a class. For example, in a tuple isa(France, country), the concept "France" is an instantiation of the concept "country". If a concept c1 occurs in a subject dependency relation and also occurs in copular verb, then the dependents of subject and copula serve as arguments. The extraction algorithm of isA relations between concepts is given in 3.4. For example, for the sentence "Kathmandu is the nation's capital", the goal is to extract the tuple isA(kathmandu, capital).

If a word occurs as a governor in a subject dependency relation and also occurs as a governor in an object dependency relation, then this word will most likely represent a relation between the dependents of the subject and object dependencies. For example, the dependency relations for the sentence "Sam is building a mall" are shown in Table 3.5.

From these dependencies, we extract the dependents (sam, mall) based on the nsubj and dobj relations. If the above method does not directly find tuples, we

Table 3.5: Dependency relations for the sentence: *"Sam is building a mall"*

Relation	**Governor**	**Dependant**
nsubj	building	Sam
aux	building	Is
det	mall	A
dobj	building	Mall

Table 3.6: Dependency relations for the sentence: *"Microsoft is headquartered in Redmond"*

Relation	**Governor**	**Dependant**
nsubjpass	headquartered	microsoft
auxpass	headquartered	Is
prep	headquartered	In
pobj	in	redmond

then look for a preposition that may come in between subject and object. If a preposition is present, it has to be glued in the relation as the meaning will change with the addition of the preposition. For example, for the sentence "Microsoft is headquartered in Redmond", the dependency relations are shown in Table 3.6.

From the above dependencies, concept tuples that is to be extracted is: $headquartered\_in(microsoft, redmond)$.

If both these methods do not extract all the tuples, and if there still exists a preposition dependency relation and an object dependency relation that has not been used for tuple extraction, then we create another tuple with only one concept as the dependent of the object relation, and relation as the combination of governor and dependent.

Some words in a sentence such as adjectives and adverbs modify or specializes concepts and relations. This information also needs to be saved in order

Table 3.7: Algorithm to extract concept relation tuples from a sentence

extract concept relation tuples
$if, \exists rel(c_1, c_2),\ where\, rel = subj\ \&\&$
$\quad \exists rel(c_1, c_3), where\, rel = obj, then$
create a tuple t, where
$\quad concept1 = c_2,$
$concept2 = c_3 and$
$\quad relation = c_1$
$elseif, \exists rel(c_1, c_2), where\, rel = subj\&\&$
$\quad \exists rel(c_1, c_3), where\, rel = prep\&\&$
$\quad \exists rel(c_3, c_4), where\, rel = obj, then$
create a tuple t, where
$\quad concept1 = c_2,$
$concept2 = c_4\, and$
$\quad relation = c_1 + c_3$
$elseif, \exists rel(c_1, c_3), where\, rel = prep\&\&$
$\quad \exists rel(c_3, c_4),\ where\, rel = obj, then$
create a tuple t, where
concept1=
$concept2 = c_4\, and$
$\quad relation = c_1 + c_3$

to have a complete transformation. These specializations are mostly done with modifiers in dependencies such appositive, possessive modifier, adverbial, quantifier modifier etc. For example, in the sentence "Microsoft is headquartered in beautiful Redmond", the word "beautiful" defines "Redmond". This is expressed as a tuple: *propertyOf(beautiful, Redmond)*. Similarly, other tuples for modifiers are extracted with the algorithm in Table 3.8.

Table 3.8: Algorithm to extract modifier tuples form a sentence

extract modifier relation tuples
$\forall rel(c_1, c_2)$, $where\, rel = appos$,
Create a tuple, $defines(c_2, c_1)$
$\forall rel(c_1, c_2), where\, rel = poss$,
Create a tuple, $belongsTo(c_1, c_2)$
$\forall rel(c_1, c_2), where\, rel = amod \cup advmod \cup quantmod$
Create a tuple, $propertyOf(c_1, c_2)$

The tuples generated during the CNF transformation of a sentence are the building blocks for other complex operations such as concept reconciliation, subsumption hierarchies, and inference.

3.6 Data

The methods we are proposing in this chapter are relevant to factual text. Our long term goal is to use all possible factual pages available in the web to create and populate our knowledge. But taking the available resources into consideration, we chose to use the pages in Wikipedia as a collection of factual pages (In, 2010). It is a multilingual, web-based freely available encyclopedia which is constructed with a collaborative effort of volunteers around the globe. Although we use Wikipedia as the dataset, we do not use specific properties of it such as infoboxes and tables which are in structured form. Our system is capable of processing any text in English with good grammatical structure, a requirement of the dependency parser we used. For the current evaluation, we only used the articles in country category. We manually created a list of countries in their corresponding capital cities and then evaluated "isa" relation for these two concepts. For comparative evaluation, we used the relations available in the labeled data created by Bunescu (Bunescu & Mooney, 2007). The data is labeled for two relations: corporate acquisitions and birthplaces. We also used the labeled data provided by Banko (Banko et al., 2008) which labeled data for the relations of inventorOf and wonAward.

3.7 Method Validation

We show a direct comparison of the result of our model compared to some popular state of the art models working in the same direction. We will test precision and accuracy of our model in six different relations with different arguments. The relations and arguments and their sizes are shown in Table 3.9

We compared the performance of our system based on precision and recall. Precision illustrates the exactness of system whereas recall shows the completeness of the system. Table 3.10 and Table 3.11 show the results compared to conditional random fields (CRF) implementation of Textrunner and multiple instance learning (MIL) method.

Our results show a significant improvement over the results of TextRunner (O-CRF), both on precision and recall. We achieve a significant improvement (more than double) on recall with an increased precision in most of the evaluated relations. One of the reasons may be because TextRunner does not attempt to model long distance relation of concepts in a sentence (Banko et al., 2009).

It only uses a window of size 6. Our method on the contrary can handle any sentence of length less than 50 and can capture relations between concepts separated by longer distance. Furthermore, it does not have a restriction that a relation has to occur in between concepts. For relations such as acquisition, birthplace, inventorOf etc, our system projects a nearly perfect precision with relatively high recall compared to TextRunner. It performs comparatively similar to multiple instance learning for acquisition and birthplace. The advantage of our system is that our system does not need to know the semantic class of participating concepts, whereas MIL has to know the type of relation beforehand. Furthermore our method is open to any relation and any concepts.

Table 3.9: Concept pairs for different relations used for evaluation

Concept1	Concept2	Size
BornIn Relation		
Franz Kafka	Prague(+)	28
Andre Agassi	Las Vegas(+)	41
Charlie Chaplin	London(+)	24
George Gershwin	New York(+)	15
Wolfgang A. Mozart	Vienna(+)	14
Marie Antoinette	Vienna(+)	39
Acquisition Relation		
Google	Youtube(+)	126
Adobe Systems	Macromedia(+)	63
Viacom	Dreamworks(+)	30
Novartis	Eon Labs(+)	30
Yahoo	Microsoft(-)	6
Pfizer	Teva(-)	9
Pfizer	Rinat Neuroscience (+)	41
InventorOf Relation		
ruth handler	Barbie Doll(+)	16
tim berners - lee	World wide web(+)	28
ted nelson	world wide web(-)	4
frank robinson	coca cola(-)	5
dean kamen	segway human transporter(+)	7
john pemberto	coca cola(+)	6
WonAward Relation		
albert einstein	nobel prize(+)	15
president truman	presidential medal of freedom(+)	3
john mccain	purple heart(+)	3
raymond damadian	nobel prize(-)	4
john steinbeck	pulitzer prize(+)	6
joseph pulitzer	pulitzer prize(-)	3
francis crick	Nobel Prize(+)	13
isA Relation		
Country	195	
Capital of country	195	

Table 3.10: Comparison of the precision of different system with respect to given relations

Relation	**Our System**	**O-CRF**	**MIL**
Acquisition	92.7%	75.6	95%
Birthplace	100%	90.6%	95%
InventorOf	100%	88%	
WonAward	100%	62.5%	
All	96.7%	75%	
isA(Country)	81.7%		
isA(Capital)	65.21%		

Table 3.11: Comparison of the recall of different system with respect to given relations

Relation	**Our System**	**O-CRF**	**MIL**
Acquisition	40.8%	19.5	44%
Birthplace	61.4%	31.1%	55%
InventorOf	56.14%	17.5%	
WonAward	65.9%	15.3%	
All	50.4%	18.4%	
isA(Country)	60.15%		
isA(Capital)	35.54%		

Here we present few limitations of our approach. For example, in the case of isA(capital) relation, our system did not perform well. Our hypothesis is that the error may be attributed to the ambiguity problem prominent in natural language. The word *capital* is used in many contexts other than denoting the *capital city of a country*. For example, *capital* could be related to *money* or could mean the *capital of a state*, etc. Hence, in order to make the system more accurate, the problem of concept disambiguation also needs to be handled which is part of our long term goal.

3.8 Conclusion

We described the process of rule based open domain information extraction from deeply parsed sentences. This method is most suited for scenarios where dataset is not large and tuples need to be extracted without any existing seed or training tuples. Due to the preprocessing complexities, scaling this method to web scale is a challenge. Generating concept tuples from a sentence is reportedly a noisy, uncertain and inconsistent process. The authenticity of these tuples can also be validated with statistical methods such as Pointwise Mutual Information(PMI) and frequency based evaluation process. We will explore these evaluation methods in the next chapter. A part of our future goal is the addition of mechanisms such as inference, synonymy detection, polysemy detection, taxonomy etc. Inference is the process of deducing a desired fact based on the reasoning of known facts. One of the properties of concepts is that they naturally form a taxonomic scheme. The discovery of this structure will help in uncovering more information that is not explicitly stated. For example: if we have tuples locatedIn(Mt. Everest, Solukhumbu) and locatedIn(Solukhumbu, Nepal) and we observe that locatedIn relation satisfies the property of transitivity, we can easily deduct tuple: locatedIn(Mt. Everest, Nepal). Another important task to make the system more accurate is to handle the problem of concept reconciliation also known as entity matching, record linkage, de-duplication; the process of identifying whether two different concept identifiers are referring to the same real world item.

Chapter 4

INDUCTION BASED CONCEPT AND RELATION EXTRACTION

"Each problem that I solved became a rule which served afterwards to solve other problems." - Rene Descartes

4.1 Overview

This chapter presents an iterative induction based relation extraction in domain agnostic environment and in large scale. Text patterns are the representations of text sequences that more than one sentence will follow to express an idea. The patterns should be general enough to cover most tuple extractions and specific enough to only extract correct tuples. In order to perform at web scale, these patterns also need to be able to work efficiently. The imperative requirement of a framework that is based on iterative pattern induction method arises with the idea of using minimal human supervision to process data in web scale in feasible time frame. A pattern is a function $p : P \rightarrow \{0, 1\}$ that returns 1 if the given context of the tuple expresses valid tuple context. Patterns are a set of constrained regular expressions. In that sense they represent a special language that is described by regular grammar and forms a non deterministic finite state automata (NFA). This automata accepts input sequence if it follows from the start state via permitted transitions to the accepting end state (Blohm, 2010).

In this chapter, we propose a system, CREATE (Concept Representation and Extraction through Heterogeneous Evidence), to extract relation tuples from large text corpora based on iterative pattern induction methodology. We will start with a single selective pattern and iteratively add tuples and patterns in corresponding collection. This method is easily expandable to any domain since it does not require any labeled data. We ensure the selectivity of pattern by filtering the patterns with statistics such as frequency and average PMI and specificity of the pattern. *CREATE works under the assumption that sentences have a pattern of* *expressing information and this pattern is followed by multiple sentences.* If we can explore these patterns in a language, we can extract tuples from all the sentences to build an automated system. One of the simplest cases of such a pattern is a sentence that only has two nouns and a verb in between. For example, for the sentence *"Google bought Youtube"*, the part-of-speech structure will be "NNP VBD NNP" and it is easy to identify two nouns as concepts and the verb as a relation between these two concepts. Thus, the tuple, *bought(Google, Youtube)* can be extracted with high confidence. The strength of this system is that it gracefully identifies such patterns without requiring any human input and expands itself with the addition of every sentence to the system. The state of art system that is closest to CREATE in terms of tuple generation is Reverb (Fader et al., 2011). The core idea of Reverb is to identify a relation and extract concepts in the immediate left and right of the relation to form a tuple. The system takes a greedy approach where it only considers concepts that are adjacent to relations. Also, it ignores information that might change context of tuple in sentence. For example, for sentence *"RSV in older children and adults causes a cold."*, Reverb extracts tuple *causes(adults, a cold)* with confidence 0.6799. This approach has two disadvantages, first; it extracts invalid tuple as it ignores complete sentence context, second; it misses correct tuple *causes(RSV, cold)* because of its greedy nature. We overcome both the disadvantages in CREATE. Although Reverb does not require training data to extract tuples, it does require labeled data to determine the confidence of a tuple. CREATE does not require labeled data other than the seed pattern at any stage of the process. With enough iterations and larger corpus, CREATE is able to extract the tuple *causes(RSV, cold)* correctly with high confidence.

Few of the properties that we exploit for filtering of tuples are as follows:

- Patterns and tuples have dual dependence. Patterns can be used to extract tuples and tuples can be used to identify patterns.

- If a tuple is generated from two different sentences using two different patterns, then the confidence of the tuple is highly increased.
- If a pattern only produces high quality tuples, then the pattern is considered to be of high confidence.
- Web is highly redundant. This redundancy can be exploited to evaluate the correctness of a tuple.

Our approach is to learn the patterns in an iterative manner as in DIPRE (Brin, 1999) and Snowball (Agichtein & Gravano, 2000). We extend the work one step further to iteratively extract tuples with open relations from large text corpora. We follow the standard step of extracting patterns based on known tuples, extracting tuples based on known patterns and evaluating and refining patterns based on inherent statistics to obtain high precision tuples and patterns.

We make the following contributions in this chapter.

- We extend and adapt pattern based tuple extraction to perform open information extraction.
- We propose a method of domain independent pattern generation.
- With the patterns generated in step 2, we propose a method of relation tuple extraction.
- We propose an effective method to refine/rank extracted tuples and patterns without human supervision.

4.2 Related Work

Patterns are the generalizations over one or more textual mentions. Most of the work done so far have developed patterns either at sentence level or part of sentence level bordered by a specified window size limit. A set of such patterns are explored in different ways to form a pattern space from textual mentions. The basic idea in pattern induction based method is to extract tuples based on existing pattern match. One of the approaches for pattern induction method is to start with very general pattern that matches a lot of tuples. This method will have a higher recall and lower precision. This approach thus adds a quality assessment layer by adding more constraints on the validity of the tuple (Brin, 1999; Pantel & Pennacchiotti, 2006). Another bottom-up method takes a greedy approach of choosing extraction patterns by imposing constraints on the pattern selection method (Mooney, 1999; Agichtein & Gravano, 2000).

4.3 Problem Definition

We formulate the problem of relation tuple extraction as a binary classification problem. Given a sentence $S = (w1; w2; ..; e1..; wj; ..r1; wk..; e2; ::; wn)$ where $e1$ and $e2$ are the entities of interest, $r1$ is the relation of interest, and $w1, w2....wj...wk$ is the context of the tuple in the sentence s, the classification function,

$$f(T(S)) = \begin{cases} 1 & \text{if } e1 \text{ and } e2 \text{ are related by } r1 \\ -1 & \text{otherwise} \end{cases} \tag{4.1}$$

Here $T(S)$ is a feature set extracted from the sentence as a context. The classification model is built based on context, independent of entities and relations. A context or a pattern of a tuple in a sentence is a 4-tuple $(left, middle\_left, middle\_right, right)$ where $left$ is the sequential list of entities and words that occur before first argument in the tuple, $middle\_left$ is the list of words that occur between first argument and relation, $middle\_right$ is the list of words that occur between relation and second argument and $right$ is the list of words that occur after second argument in the sentence unless another relation is detected.

The classification function $f(T(S)) = 1$ if the pattern of the tuple T in the sentence S exists in pattern database. The degree of similarity of the context of probable tuple is greater than threshold similarity with one of the contexts existing in context-base.

4.4 CREATE Tuple Pattern Extraction Methodology

Given a set of documents containing sentences, our goal is to extract relation tuples with highest recall and precision. As explained earlier, our system is designed to utilize the dual dependence of tuple with pattern and pattern with tuple. As a starting point, we use a seed pattern $p = (\phi, \phi, \phi, \phi)$ that will generate tuples from text corpus. These tuples are then used to generate extraction patterns which in turn generate more tuples, an approach similar to Snowball. At this stage, all the extracted tuples and patterns in the process are not guaranteed to be correct. A good tuple should be syntactically and semantically correct as well as articulate, autonomous and informative. Similarly, a good pattern should achieve a good balance between two competitive criteria; *specificity* and *coverage*. *Specificity* means the pattern is able to identify high-quality relation tuples; while *coverage* means the pattern can identify a statistically non-trivial number of good relation tuples. Hence, in the process, we have a self evaluating system which evaluates and filters out invalid tuples and patterns based on their statistical properties. The overall system can be broken down into several modules, each of which perform an isolated task such as concept extraction, relation extraction, probable tuple generation, tuple verification etc. The system architecture of the overall system has been depicted in figure 4.1 and the algorithm is shown in Table 4.1. The sub-modules are explained in detail in the subsequent sub-sections.

1. **Feature**:We consider lexical and shallow parse information as features for relation extraction. Lexical and shallow NLP techniques are robust and fast enough for a problem like ours where extraction needs to be performed at web

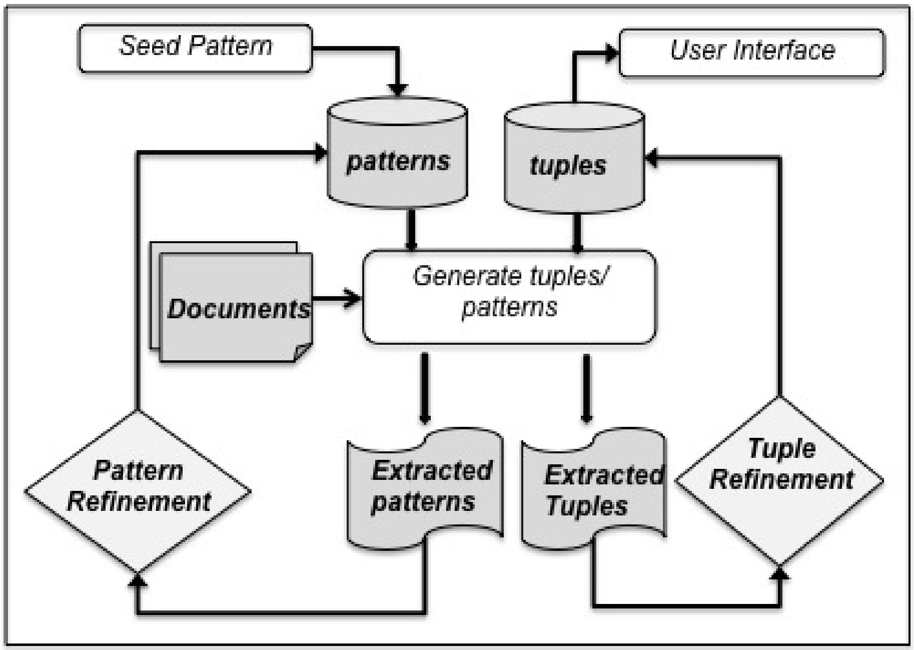

Fig. 4.1: Overall System Architecture

scale. Although, our concept extraction module can be easily replaced with named entity extractor, we primarily use part-of-speech tagging and chunking results for concept/relation extraction. All the sentences in our data sets are parsed using an OpenNLP (Baldridge, Morton, & Bierner, 2004) part-of-speech tagger.

2. **Seed Pattern**: We start with a fairly general and yet very strict pattern that will extract tuples from a sentence. The seed pattern, $p_s = \{\phi, \phi, \phi, \phi\}$ meaning there is an empty left context, empty middle left context, empty middle right context and empty right context. As an example, let us consider a sentence *"Temperature is ultimately regulated in the hypothalamus"*, our process extracts two concepts "Temperature" and "the hypothalamus" and relation "is ultimately regulated in". The left context (context before concept 1) in this case is empty, middle left context (context between concept 1 and relation) is also empty and similarly, middle right and right contexts are empty. This is a fairly specific pattern for a tuple to be valid and moreover,

this pattern is domain independent and can be applied to any domain for english language. We have a running example showing the steps in Table 4.2.

3. **Concept Extraction Module**: We extract concepts in the sentence based on noun phrases. We remove starting and trailing stopwords in noun phrases. If noun phrases contain conjunction, we break down noun phrase into two concepts.

4. **Relation Extraction Module**: To extract relations, we extract the longest sequence of words such that it starts with verb or is a sequence of noun, adjective, adverb, pronoun and determiner or a sequence of preposition, particle and infinitve marker. If any pair of matches are adjacent or overlap in a sentence, we merge them to a single relation. This method has been proven to be effective in (Fader et al., 2011).

5. **Probable Tuple Extraction**: For each relation $r \in R$ and for every combination of $c_i \, and \, c_j \in C$, such that c_i occurs before r and no other relation occurs between $c_i \, and \, r$, and c_j occurs after r and no other relation occurs between $c_j \, and \, r$ in the sentence, we create a probable tuple $t = (c_i, r, c_j)$.

6. **Tuple Pattern Extraction**: For each tuple $t = (c_i, r, c_j)$ in sentence s, we extract the sequence of words in sentence that occurs between beginning of sentence and concept c_i. If a relation occurs before c_i, we start with the end of closest relation. This is the left context. Similarly, we extract middle_left context as the sequence of words between c_i and relation r. Middle_right context is the sequence of words between relation r and c_j. Right context is the sequence of words between c_j and either another relation r_p (if exists) or end of the sentence. We experiment with three types of patterns, first: purely lexical(only use lexicons for pattern generation), second: purely syntactic (only use part of speech tags for pattern generation) and third: mixed

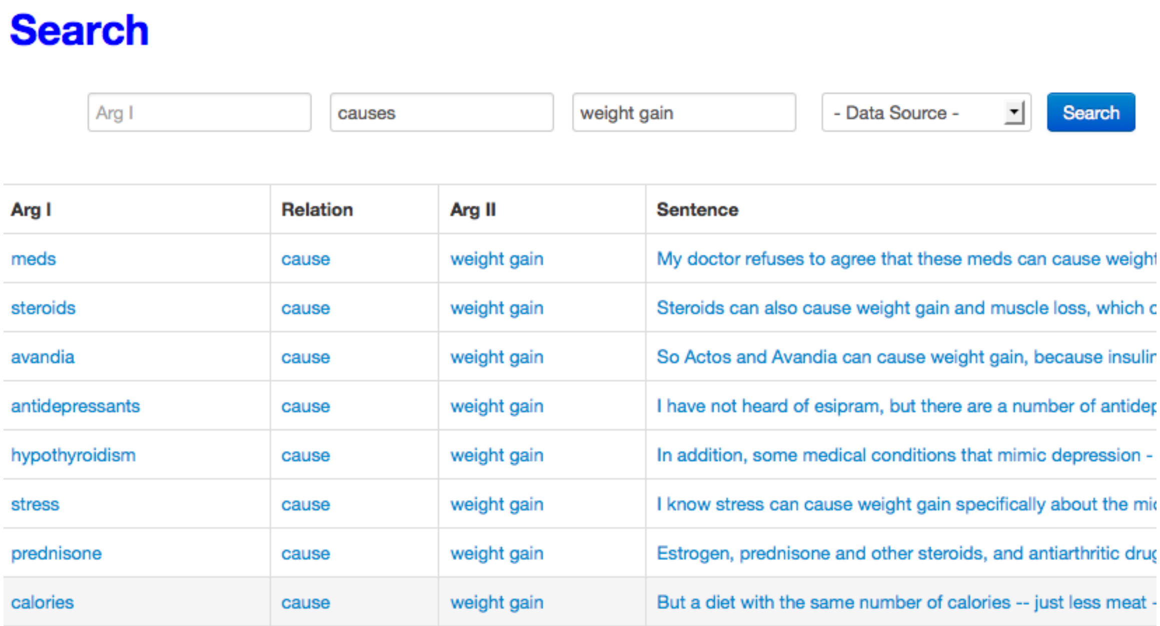

Fig. 4.2: Concept based Search User Interface

pattern(a combination of lexicons and part of speech tags. For mixed pattern, we replace all nouns, verbs, adjectives and adverbs with their part of speech tags and leave preposition, particle and other words to use lexicons.

7. **Iteration**: Our system is an iterative process and gets better qualitatively and quantitatively with each iteration. The number of iteration is highly dependent on the application of interest, pattern database size, size of corpus and time sensitivity of the system. We experimented on a sampled data to see the convergence of the algorithm. We also iterated over a large corpus to see the $\vec{\beta}$ and tuples. As the extraction algorithm is based in active learning methodology, the system can perform quite well with iteration count as small as 2 in large corpus.

4.5 Tuple Refinement

4.5.1 Tuple and Pattern Filtering

We employ a holistic approach for concepts and relations extraction that enforces coherence in relations and concepts in tuples. To ensure validity of

Table 4.1: Iterative Pattern Induction Algorithm

Algorithm 1 Iterative Pattern Induction

Input: $Pattern, P = \{seed\_pattern\}$,

$Tuples, T = \{\phi\}$

$Sentences, S = \{s_1, s_2, ...s_n\}$

Output: $Patterns, P = \{p_1, p_2, ...p_x\}$,

$Tuples, T = \{t_1, t_2, t_3......t_y\}$

for every $S_i \in S$ **do**

$C_{prob} = \{c_1, c_2, ..c_j\} \leftarrow extractConcepts(S_i)$

$R_{prob} = \{r_1, r_2...r_h\} \leftarrow extractRelations(S_i)$

$p_{sent} = replaceConceptsRelations(C_{prob}, R_{prob})$

$T_{prob} = \{t_1, ..t_u\} \leftarrow extractProbableTuples(C_{prob}, R_{prob})$

end for

for every $t_j \in T_{prob}$ **do**

$pattern, p_i = extractPatternFor(S_i, p_s)$

if $p_i \in P$ && $t_j \notin T$

$T.add(t_j), P.update(p_i)$

else if $p_i \notin P$ && $t_j \in T$

$P.add(p_i), T.update(t_j)$

else if $p_i \in P$ && $t_j \in T$

$P.update(p_i), T.update(t_j)$

end if

end for

extracted tuples, we select patterns and tuples that occur more than α (3 in our experiments) and β (2 for medical and 1 for Wikipedia for our experiments) times respectively. The total frequency of a pattern p in a relation r is defined as the sum

Table 4.2: Running Example of Tuple and Pattern Extraction

Parameter	**Value**
seed pattern	$(\phi, \phi, \phi, .)$
sentence	Sunscreen may also cause drying of skin.
concepts	Concept1=Sunscreen, Concept2=skin
relations	relation=may also cause drying of
sentence pattern	Concept1 relation Concept2.
probable tuple	may_also_cause_drying_of(sunscreen, skin)

of the frequencies of p in all entity pairs that have relation r. We define confidence of a tuple as follows:

$$Conf(t) = \frac{\sum_{p \in P_t} f(p_i)}{f(p_{max_t}) log(N)} \tag{4.2}$$

where $f(p_i)$ is the frequency of pattern p_i for relation r such that tuple t also has relation r. Here, $f(p_{max_t})$ is the frequency of pattern that has maximum frequency for relation r and N is the total number of distinct patterns that match tuple t. Note here that confidence conf(t) can be greater than 1 depending on the number of patterns that extract tuple t.

4.5.2 Tuple relevance

Traditional vector space model based relevance cannot be applied to concept based relevance paradigm. Hence, we employ PMI based relevance for tuple retrieval. If e1 is the query entity for which search is executed, then the relevance of a tuple is calculated in terms of PMI between query entity e1 and second argument in tuple that contains e1 as first argument. PMI between entities e1 and e2 is defined as

$$PMI(e_1, e_2) = log \frac{P(e_1, e_2)}{P(e_1, e)P(e_2, e)} = logN \frac{n_{12}}{n_1.n_2} \tag{4.3}$$

$$NPMI(e_1, e_2) = \frac{PMI(e_1, e_2)}{-logP(e_1, e_2)} \tag{4.4}$$

where N: the total number of tuples in the corpus, $P(e_1, e_2) = n_{12}/N$=the number of sentences containing tuples that have e_1 and e_2 as arguments,
$P(e_1, e) = n_1/N$: the probability that the entity e_1 co-occurs with entity e in tuples,
$P(e_2, e) = n_2/N$: the probability that the entity e_2 co-occurs with entity e in tuples.

4.6 Prototype and Experiments

4.6.1 System Prototype

We built the system prototype CREATE based on the process explained in this chapter for two datasets, namely; wikipedia and medical sites. We crawled top medical information sites and collected sentences talking about medicine. The list of medical sites crawled are mentioned in Appendix 1. Since medical sites contain information about the same domain and have higher redundancy, this method was more effective in such environment. Wikipedia articles talk about different real life concepts and hence have little redundancy. For wikipedia, we used named entities tagged in a different process to obtain concepts (Atserias, Zaragoza, Ciaramita, & Attardi, 2008). CREATE provides a tuple searching interface and a concept-graph based navigation system. We demonstrate the usefulness of the system with medical information and evaluate against few relations in Wikipedia.

4.6.2 Case Study

A user Alice wakes up early in the morning, has fever and worries if she has flu symptoms. She goes to a major search engine and tries to find flu symptoms. She reads through few web pages and learns about few flu symptoms. What she would ideally want is that when she issues a query *"what are flu symptoms ?"* she would expect a list of flu symptoms. We issued a query "flu symptoms" as first

Table 4.3: Results returned by CREATE Tuple Search System

flu symptoms include
high fever
stuffy nose
chills
body aches
runny nose
headache
muscle aches
fatigue

argument and "include" as relation. We obtained the result shown in Table 4.3 based on the query in CREATE system. Figure 4.2 shows a snapshot of the prototype for medical data for another example.

4.6.3 Comparison with Open Information Extraction Systems

We compared the result of our system with other systems such as Reverb, TextRunner and WOE. For evaluation purpose, we used the test set of 500 sentences used in Reverb system evaluation (Fader et al., 2011). The figures shows the quantitative comparison of our system compared to reverb and woe. It has to be noted, however that this result does not evaluate the iterative process of create. The distinctive advantage of create is seen when applied to a relatively larger corpus where the system is applied iteratively.

Figure 4.3 and 4.4 show the effect of iteration on tuples and patterns with the CREATE algorithm. It shows that in initial iterations, there is a rapid increase in number of patterns and tuples, which starts to converge with higher iterations. For proof of concept, we experimented with a sample data that we created with

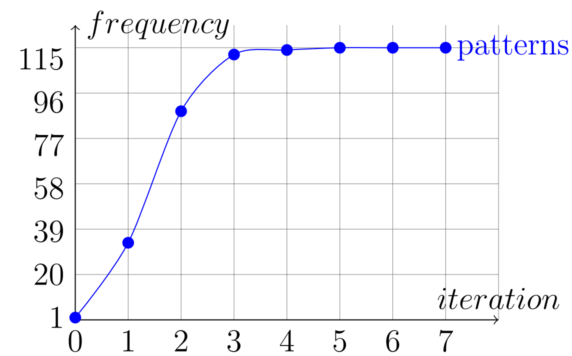

Fig. 4.3: Effect of Iteration on Number of Patterns

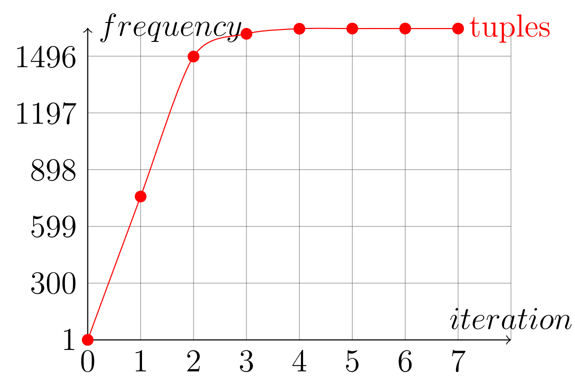

Fig. 4.4: Effect of Iteration on Number of Tuples

medical sentences. It shows that tuple and pattern generation converges in 5 iterations.

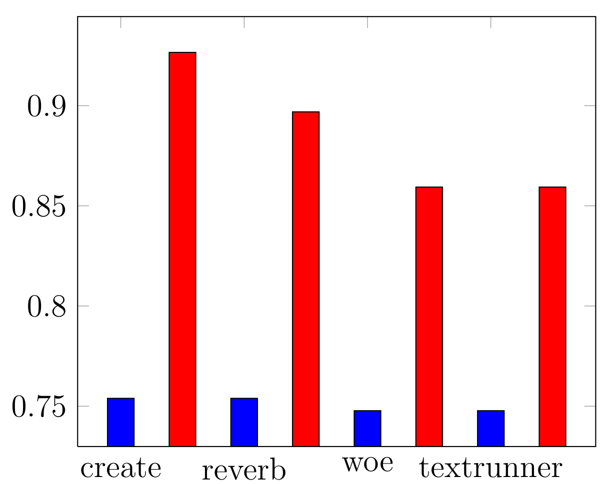

Fig. 4.5: Performance comparison of CREATE with Reverb, WOE and TextRunner

Figure 4.5 shows the comparison of CREATE with Reverb, WOE and TextRunner. We see improved recall at around 92% and precision around 75% for CREATE which outperforms all other systems. Similarly, figure 4.6 shows the effect of iteration on the performance of CREATE. We see the same effect of rapid increase in performance in initial iterations and then it gets stabilized after few iterations.

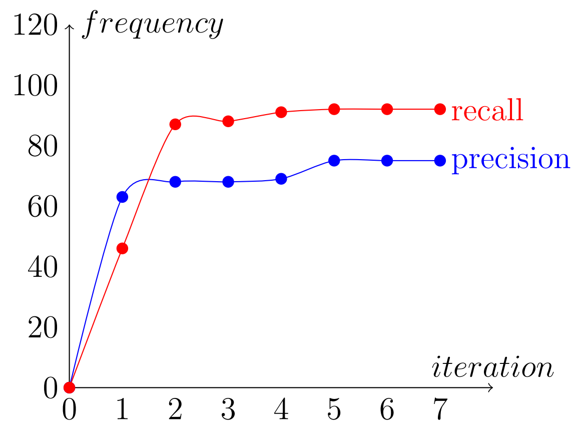

Fig. 4.6: Effect of Iteration on Tuple Extraction Performance with confidence 0.6

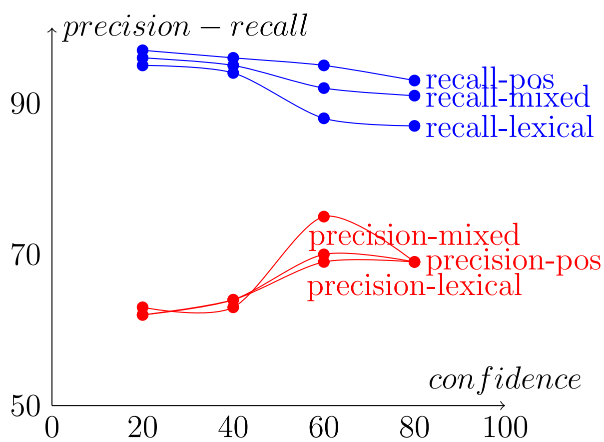

Fig. 4.7: Precision-Recall variance with Confidence

We also experimented with the performance based on different patterns. Figure 4.7 shows that recall for POS pattern is the highest but the precision is highest with mixed pattern.

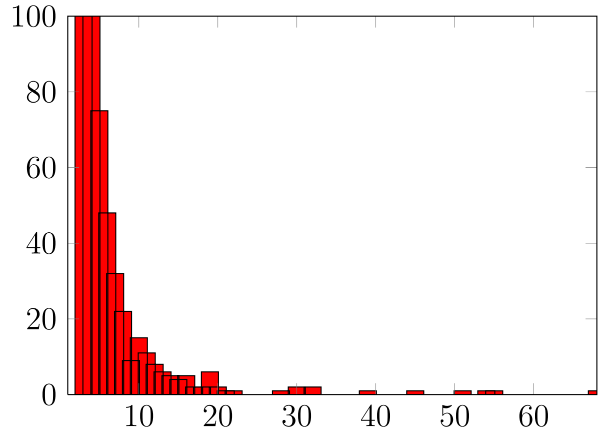

Fig. 4.8: Distribution of Pattern Frequency

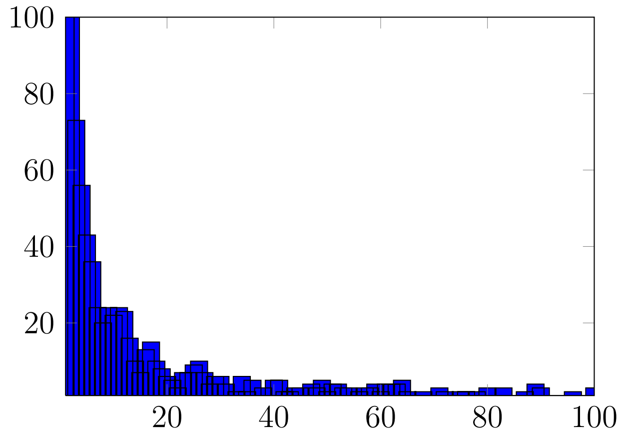

Fig. 4.9: Distribution of Tuple Frequency

Figures 4.8 and 4.9 show the distribution of patterns and tuples based on their frequency. As expected, there were few tuples and patterns with high frequency and many tuples and patterns with low frequency or single occurrence.

4.6.4 Wikipedia Tuple Extraction

We used Semantically Annotated Snapshot of the English Wikipedia to extract relation tuples as the first large dataset (Atserias et al., 2008). The SW1 corpus is a snapshot of the English Wikipedia dated from 2006-11-04 processed with a number of publicly available NLP tools. We chose to use this data as it has been processed and has information on shallow parsing such as POS tags and named entities on seven categories. To demonstrate the interchangeability of concept

Table 4.4: Data statistics for Wikipedia

Relation	**Gold Data**	**Create (total/ correct)**	**Precision**	**Recall**
bornIn(x,Atlanta)	440	341/303	88.8	68.8
bornIn(x,Zurich)	108	87/75	86.23	69.4
graduatedFrom(x,Stanford)	456	403/345	85.6	75.6
graduatedFrom(x,Princeton)	582	464/385	82.9	66.1
presidentOf(x,United States)	44	65/39	60	88.86

extraction module, we used named entities as concepts for relation extraction. We then generated tuples from data. Since it is not possible to evaluate all the relation tuples extracted from Wikipedia, we performed samples evaluation of the system for few sampled relations and tuples. We compared the performance of our system based on precision and recall compared to DbPedia. The evaluation in terms of precision and recall is shown in Table 4.4. Precision and recall are given by the following equations

$$precision = \frac{|(correct\ docs) \bigcap (retrieved\ docs)|}{|(retrieved\ docs)|} \tag{4.5}$$

$$recall = \frac{|(correct\ docs) \bigcap (retrieved\ docs)|}{|(relevant\ docs)|} \tag{4.6}$$

4.7 Conclusion

This chapter introduces a novel iterative pattern induction based relation extraction method that is purely unsupervised and can achieve comparable results to semi-supervised methods. This pattern induction method works on few

Data	Wikipedia	Medical
Document count	1431178	348284
Sentence count	36117170	4049238
Tuple count	6945440	1535293
Relation count	1847116	706359
Relation with freq > 9	1131	1865
Concept count	2673192	106263
Extraction latency (for single iteration)	5 hrs	2hrs

assumptions of text corpus. It assumes that there are one or more uniform ways in which instances of a target relation are mentioned in text that distinguishes them from other mentions. This uniformity can be observed by looking at the contexts of a limited set of mentions. Secondly, it assumes that tuples that are derived during one iteration should improve the model for the next iteration. Tuples are supposed to be mentioned in multiple contexts. The idea behind the system is that the system should be able to work autonomously and learn from web without requiring any supervision. We qualitatively and quantitatively demonstrated the effectiveness and usefulness of iterative pattern induction framework. With increasing data being added to web, the value and importance of systems such as CREATE is ever increasing. We have demonstrated the prospects of relation extraction systems. At the same time, we also need to be aware of the challenges that need to be solved before we can realize a fully functional machine reading system.

Chapter 5

CONCEPT REPRESENTATION AND KNOWLEDGE MANAGEMENT

"Reality exists in the human mind, and nowhere else." George Orwell

5.1 Overview

The concept of "Linked Data" as coined by Tim Berners Lee, refers to the process of exposing and connecting structured data on the web. In a broader context, the goal is to be able to query the web like a structured relational database. Structured information obtained during the knowledge discovery process such as open domain information extraction explained in previous chapters need to have a well defined knowledge representation process in order to make the best of semantic processing. Knowledge discovery/concept extraction processes provide structured data that can be combined and adopted to form a loosely coupled semantic hierarchical structure. The idea here is to organize relation tuples in order to realize a web of concepts. Semantic Web along the same line, envisions a complete formalized and explicit semantic information represented in terms of strict ontologies. In this chapter, we will discuss the process of organization of relational tuples obtained during open domain information extraction process into a lose hierarchical form.

Semantics by nature is informal and implicit. Formal ontologies are one of the ways to depict explicit semantic information and formalize them to a definite structure. Thus to envision a complete machine reading system, we adopt the idea of ontology based knowledge representation. As opposed to manually created ontologies, we work towards automatically created ontology based on extracted concepts and relations. In addition, in this chapter, we will also explain the details of the persistence design and implementation for our concept based knowledge management model along with the indexing details for user interface.

Ontology based knowledge management mainly relies on the resource description framework (RDF) to save explicit semantic information. RDF is based upon the idea of making statements about resources (in particular web resources) in the form of subject-predicate-object expressions.These expressions are known as triples in RDF terminology. The subject denotes the resource, and the predicate denotes traits or aspects of the resource and expresses a relationship between the subject and the object. In semantic web world, the subject, object and predicate of a triple all are URIs such that each identify a resource. We have similar notion of triples in our system although we name it as a relation tuple mainly because the tuple defines a relation between two entities. A collection of relation tuples will form a labeled directed multigraph as shown in figure 5.4. However, since we are advocating for an automatically generated ontological structure, we don't have a completely disambiguated data, hence we do not have a URI based identifier for each concept.

Maintaining and organizing these explicit semantic tuples so that it is helpful in processes such as inference and query processing is a difficult and time consuming process specially as tuple count increases. The system should be consistent, self-improving in terms of confidence and should be able to detect redundancy and anomalies. Previous work on semantics based knowledge management has mainly been concentrated in XML-based knowledge base creation. We will talk about ontology based knowledge management in the related work section. Semantic web standards are not directly applicable to noisy and possibly conflicting data obtained from open domain information extraction process. We will thus need a system resilient to such data. We introduce the concept of confidence for each tuple. The higher the confidence, chances are that the tuple is correct. This also means that we could have tuples that are contradicting to each other due to the noise in data and in the system.

5.2 Related Work

Most of the efforts towards standardizing semantic data use XML based languages. The metadata standard RDF (Resource Description Framework) and RDF schema language (RDFS) are used to express information and define ontology respectively. Some of the earlier works in ontology based information retrieval are Ontobroker (Naing, Lim, & Goh, 2010) and SHOE (Heflin, Hendler, & Luke, 1999). SHOE brought up the idea of using ontologies to annotate information in web as HTML pages are annotated. Swoogle, on the other hand, is a crawler-based indexing and retrieval system for the Semantic Web (Ding et al., 2004). It searches over RDF documents and indexes related documents with character N-Gram or URI references. Swoogle also generates a ranking metric to predict the popularity of a semantic web document. As opposed to Swoogle, which is still based on keyword based searching, SWSE offers an entity based search with keywords, type restricted keyword and entity browsing based on RDF (Harth, Hogan, Umbrich, & Decker, 2008). WATSON is a complete infrastructure that automatically collects, analyses and indexes ontologies and semantic data available (d'Aquin et al., 2007). It provides keyword based search over semantic web documents as well as entity based search.

5.3 Application Oriented Indexing and Querying

A completely formalized ontology based semantic web is out of the scope of this dissertation. Our work is towards a more lightweight ontological structure that can be derived from the tuples extracted in open domain information extraction process. We also offer a keyword based user interaction interface that meets the requirement of efficiency and precision. The system is geared towards presenting a direct answer to queries instead of recommending a selection of related documents. The system should also be capable of integrating heterogeneous data from multiple sources seamlessly. It indexes relation tuples instead of whole HTML documents.

This principally allows more advanced querying and browsing experience with aggregated results generated from multiple documents.

This entails that the query entered by user is matched against different components of a tuple, namely; subject, relation and object. The interface also allows to explicitly mention that the query should be searched against a particular component or a combination of components. We categorize queries entered to CREATE:FACT in two types as follows.

1. *Naive query*: The user issues a query (combination of one or more words) and based on this every component; subject, relation and object of a tuple is searched in inverted index. Example of a naive query could be an entity or a relation such as "flu", "Brad Pitt" etc. These kind of queries will most likely lead to navigational browsing as the user's intention is to explore about it.

2. *Informed Query*: The user is well aware of the subject, relation or object he/she will be looking for and hence issues the query according to that. Example of such queries could be:
   subject:flu,relation:causes, object:?
   subject:?,relation:was born in, object:Atlanta

Figure 5.1 depicts the overall crawling/indexing structure employed by CREATE. Web pages are crawled from the web, preprocessed and broken down to sentences. Tuples are then extracted iteratively from these sentences and persisted in structured database. We will talk about the database schema in detail in next section. Tuples are then indexed in components of subject, relation and object in Lucene inverted index. Queries are thus matched against this index and based on this match, sentences and thus documents are referenced. The results are presented in aggregated form in tuples and corresponding sentences structure. A snapshot of the result structure was presented in the previous chapter in Figure 4.2

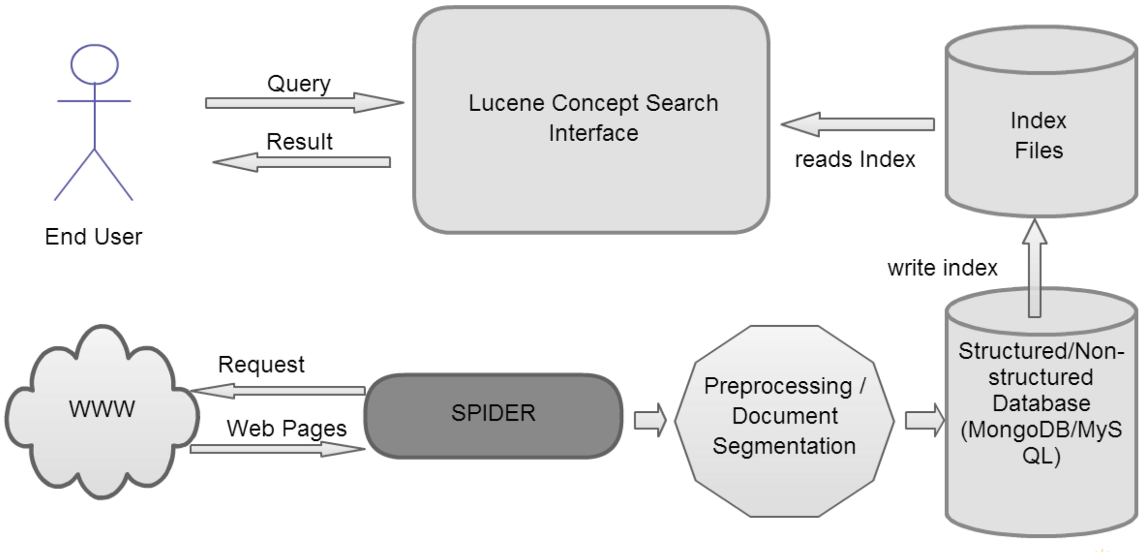

Fig. 5.1: A general view of Index-based User Interaction, Data Collection and Processing Model

5.4 Persistence Layer Design and Implementation

In earlier years, data based on its structure could be classified either as fully structured data or fully natural text data. Structured data utilize persistence systems such as relational databases (MySql, Oracle) whereas natural text utilize inverted index structures such as Lucene for fast and efficient retrieval. In recent years, persistence systems that support semi-structured data storage have been popular, few examples being XML, JSON stored in NoSql structures (mainly key-value storage). Information extraction in general is the bridge between structured and unstructured information. Our task here is to design and implement overall objective of improved performance, data exposure and data normalization. The type of data that we need to save for a complete visualization of CREATE has both the structured components and unstructured components. Hence, we use MySql for the structured component and Lucene for the unstructured component of data. User interaction/input query occurs in the indexed structure in lucene. Lucene then has references to the structured part of data such as sentence, document, tuple, concept, relation etc. The design and implementation of search based system is partly influenced by traditional search architecture in the sense that tuple components are persisted in inverted index structure. However, principally, this is more advanced than performing keyword based search in documents. We have discussed about the query structure and index counterpart in earlier section. Here, we will go into detail in the structured data counterpart. Each web-page with a unique uri will indicate a document. A document will thus have a uri, an id, a title and a list of sentences. A sentence will have a unique id, collection of documents in which this sentence exists, its frequency in multiple documents and a list of tuples that is extracted from the sentence. A tuple contains an id, collection of sentences from which this tuple is extracted, its frequency based on multiple sentences, a subject(governor), a relation and an object(dependant). Both subject and object are either concepts or instances of concepts. A concept will have a list of wordIds, frequency of itself in unique tuples and an id. Similarly, a relation will have a list of wordIds, frequency of itself in unique tuples and an Id.

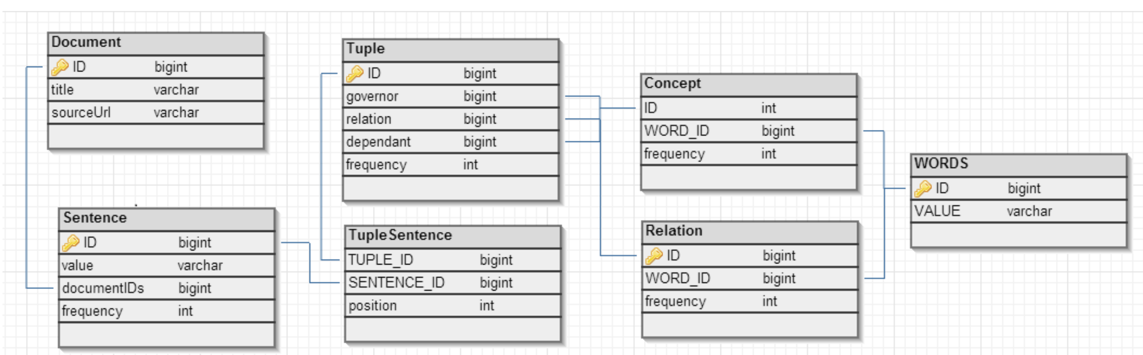

Fig. 5.2: A general view of Heterogeneous Data types and their inter-relationships for Concept based Data Representation

The overall relationships between different components of the system is depicted in figure 5.2.

5.5 Ontology Induction

As discussed earlier, building a complete infrastructure for semantic web is beyond the scope of this dissertation. Our goal here is to demonstrate the feasibility of ontology induction based on tuples extracted with open domain information extraction. Here, we propose a more relaxed system where every tuple does not need to be in agreement with all other tuples. The knowledge extracted by open domain information extraction systems in the form of relational tuples are mere triples with a relation connecting either concepts or entities. The question thus to be answered is how these tuples fit with other tuples. We present a proof of concept to verify, organize tuples in a hierarchical manner is possible. We start wih an ontology definition. *An ontology for a domain D can be defined as a specification of a conceptualization of D describing D.* The ontology will consist of following components.

- A list of concepts important for domain D
- A list of attributes describing the concepts
- A list of taxonomical (hierarchical) relationships among these concepts
- A list of (non-hierarchical) semantical relationships among these concepts

As an example, we show each concepts of ontology corresponding to medical domain in the following section.

- Concepts: disease, symptom, chemical, body-part etc.
- Attributes: disease *has* symptoms, chemical *is used in* medicine, etc.
- Taxonomical relations:
    - Flu is a disease; cymbalta is a medicine; surgeon is doctor, etc.

- Non-hierarchical relations:
    - doctor treats patient; flu causes headache; cymbalta has side-effects, etc.

We will describe the process of fuzzy ontology induction with an example in medical domain and show corresponding results in the corresponding section.

5.5.1 Extraction of concepts

For a given domain D, there are groups of entities falling under different major categories known as concepts. To extract such concepts from collection of tuples, we choose hierarchical tuples. We classify hierarchical tuples based on relations such as isa, are, called etc. We will discuss about hierarchical tuples identification in the following section. We explain the details of extracting concepts for a domain in the following algorithm.

Table 5.1: Domain Concepts Identification Algorithm

Domain Concepts Identification
Input: $Tuples, T = \{t_1, t_2, t_3......t_y\}$
Output: $isa\ Relation\ cluster,\ R = \{r_1, r_2, ...r_x\}$
1: $\forall tuple\ t_j\ such\ that\ t_j = r_i(c_1, c_2)\ where\ r_i = "isA"$ or its synonyms
2: **if** $T_{sub} = \{t_1, t_2, ...t_z\}\ such\ that\ \forall t_x = r_i(c_1, c_2),\ r_i\ and\ c_2$ are same for all tuples
3: $C = \{c_1, c_2...c_h\} \leftarrow c_2$
4: $\forall tuple\ t_j\ such\ that\ t_j = r_i(c_2, c_1)\ where\ r_i = inverse\ of\ "isA"$ or its synonyms
5: $C = \{c_1, c_2...c_h\} \leftarrow c_1$

Table 5.1 shows the algorithm to extract concepts of a domain. A sample concept list for medical domain is shown in Table 5.2.

Table 5.2: Sample concepts for medical domain

Medical Concepts
technique, birth, person, sterile, reason, test, treatment, concern, diagnosis, physician, people, cure, enzyme, story, time, nature, protein, chemical, prevention, doctors, place, brain, option, information, inflammation, problems, heart, drugs, death, answer, number, blood, health, depression, goal, conditions, bit, system, combination, adults, lot, years, small, ways, symptom, plant, expert, complex, common, choice, simple, diabetes, chronic, pain, excellent, decision, kind, effective, clear, substance, life, patient, patients, easy, times, baby, situation, ability, procedure, water, medication, safe, trigger, cases, anxiety, ms, factor, fault, research, process, point, symptoms, fact, challenge, days, bacteria, brand, injury, method, part, cancer, hormone, unsafe, body, individual, diet, change, tool, exception, study, good, type, long, studies, surgery, skin, doctor, rare, priority, woman, family, age, medicine, key, work, condition, researchers, stress, white, effort, possibility, kids, fun, cost, sign, mother

We also show some sample instances of medical concepts extracted in the previous step along with their counts for each concept in Table 5.3

5.5.2 Extraction of attributes of concepts

For the extraction of attributes of concepts, we employ the logic proposed by (Banko et al., 2009). For each concept identified in previous step, we use pointwise mutual information (PMI) to compute ranking of relations relative to the concept. PMI, defined here as

$$PMI(R, C) = \frac{Count(e_1, \in, C, R, *)}{Count(*, R, *)} \tag{5.1}$$

Table 5.3: Sample concepts and their respective instances along with counts

condition	622	tpsa, long qt syndrome, priapism, gynecomastia, autoimmune disorder, anoxia, hairy tongue, anemia, sickle cell anemia, lymphangiomatosis, congenital lactose intolerance
challenge	61	technologies, gf, large piece of bone missing, short term memory, preterm birth, levels of pain
doctor	61	gastroenterologist, primary healthcare physician, rheumatologist, dr jerald winakur, otolaryngologist, dermatologist, geriatrician, orthopedic sugeon, neurologist, cardiovascular surgeon
pain	41	headache, pelvic pain, heartburn, fatigue, tylenol, nerve pain, wrist pain, angina, earache, stiff neck
surgery	31	splenectomy, tumor, capsular contracture, bariatric surgery, gastric bypass, laparoscopy, radiation therapy, primary treatment,vasectomy,myomectomy, rhinoplasty
disease	74	trench fever, multiple sclerosis, lumbago, emphysema, measles, rheumatoid arthritis, cardiomyopathy, alzheimer, pellagra, colon cancer, silicosis, hepatitis, histoplasmosis
test	45	upper endoscopy, mri, emg, electroencephalogram, echocardiogram, ecg, elisa, spirometry, magnetic resonance imaging
drug	43	buphenyl, oxycodone, methadone, metformin, prednisone, baclofen, abilify, warfarin, naltrexone kiren, gleevec, synthroid, antihistamine, methotrexate, cymbalta, primaquine, clozaril

Here PMI measures the association between the relation R and the instances of concept C.

We show some instances of concept Cancer along with corresponding NPMI

Table 5.4: Sample concepts and their corresponding attributes discovered

Concept	Attribute	NPMI
Condition	need	0.601496346969052
	may include	0.10796014235676903
	may increase	0.10575983922898204
	caused by	0.1389892680245651
	impaired	0.10575983922898204
	including	0.13007994354075442
	treated with	0.10244955050738917
	may be	0.10435226114521522
	characterized by	0.12876485559979023
	include	0.12030990432635387
	associated with	0.13523082227171063
Challenge	need	0.42354416937544254
	developing	0.102443720124156
	having	0.07819879321534781
	has	0.07411554521276172
	is	0.0807133447039512
disease	causes	0.120067943

Table 5.5: Sample instances for concept *Cancer* along with their NPMI

Cancer Instance	**Normalized PMI**	
tumor	0.2780941674435814	337
colon cancer	0.5002160741703584	181
dcis	0.28178812424640765	71
breast cancer	0.7118487825819142	914
lymphoma	0.253470088046048	182
mesothelioma	0.2537577478933515	57
cervical cancer	0.5071043664207114	115
chemo	0.23755291987559643	229
cure	0.16306736845283623	393
woman	0.12474704760763099	205
mammogram	0.2251990524297257	79
colorectal cancer	0.5112498224351137	144
lump	0.2151616120770966	108
diagnosis	0.14709539843005145	1023
sarcoma	0.24617296560869875	61
breast	0.6229914448543666	19641
stage	0.36133775692353753	2571
leukemia	0.23438793833392388	198
prostate cancer	0.5668880471355382	294
breast pain	0.2081533635457773	59
prostate	0.4966882848257278	4893
biopsy	0.23507900217534822	203

in Table 5.5. The values clearly show that higher NPMI values suggest higher confidence of correctness.

5.5.3 Synonymous Relations Extraction

We employ a simple algorithm to detect synonymous relations in our tuple corpus. The idea is that if multiple relations occur between same pair of concepts in a sizable number of tuples, chances are that the relations are synonymous to each other.

Table 5.6: Synonymous Relations Identification Algorithm

Synonymous Relations Identification
Input: $Tuples, T = \{t_1, t_2, t_3......t_y\}$
Output: *isa Relation cluster*, $R = \{r_1, r_2, ...r_x\}$
1: $\forall r_i \in R$ *where* $r_i.frequency > 4$
2: $\forall$*tuple* t_j *such that* $t_j = r_i(c_1, c_2)$
where $pmi(c_1, c_2) > 0.2$ && $t_j.frequency > 2$
3: $R_{prob} = \{r_1, r_2...r_h\} \leftarrow allrelationsbetween(c_1, c_2))$
4: $R^{-1}_{prob} = \{r_1, r_2...r_h\} \leftarrow allrelationsbetween(c_2, c_1))$
5: $\forall r_k \in R_{prob}$ *and* R^{-1}_{prob} **do**
6: **if** $r_k.frequency > 0.5 * r_i.frequency$
7: $R_{syn} = \{r_1, r_2...r_h\} \leftarrow r_k$
8: $R^{-1}_{syn} = \{r_1, r_2...r_h\} \leftarrow r^{-1}_k$

The algorithm is explained in Table 5.6. Table 5.7 shows some sample isa relation synonyms along with inverse synonyms.

5.5.4 Extraction of Hierarchical Relationships

To extract all the hierarchical relations, we start with the obvious isa and then bootstrap other hierarchical relations. To extract other hierarchical relations,

Table 5.7: Sample relation groupings representing synonymous and inverse synonymous relations to *"isA"* relation.

Synonyms	**Frequency**	**Inverse synonyms**	**Frequency**
are	672	called	857
include	112	associated with	178
will be	111	found to be	118
having	101	may be	146
can be	369	having	175
called	100	was	305
may be	383		
was	968		

we employ the same algorithm used for synonymous relations extraction. The algorithm is already explained in section 5.5.3. We present a miniature hierarchical structure induced from tuples extracted with open domain information extraction in Figure 5.3. With such induction of hierarchy, common-sense inference is easily deducible. As an example, with the structure in figure, we can easily induce a new tuple *isA(brain tumor, disease)*.

5.5.5 Extraction of non-hierarchical relationships

These are the remaining grounded tuples that express relationship between two concrete entities. These tuples carry explicit fact about real world entities. The non-hierarchical tuples form a multigraph of entities connected by different relationship as depicted in Figure 5.4. Facts are represented in terms of these tuples whereas structural information is represented in the form of hierarchical tuples.

5.6 Conclusion

Like most other knowledge formalisms, in this chapter, we make a statement to define classes and their instances of real life entities. Entities are related to one or

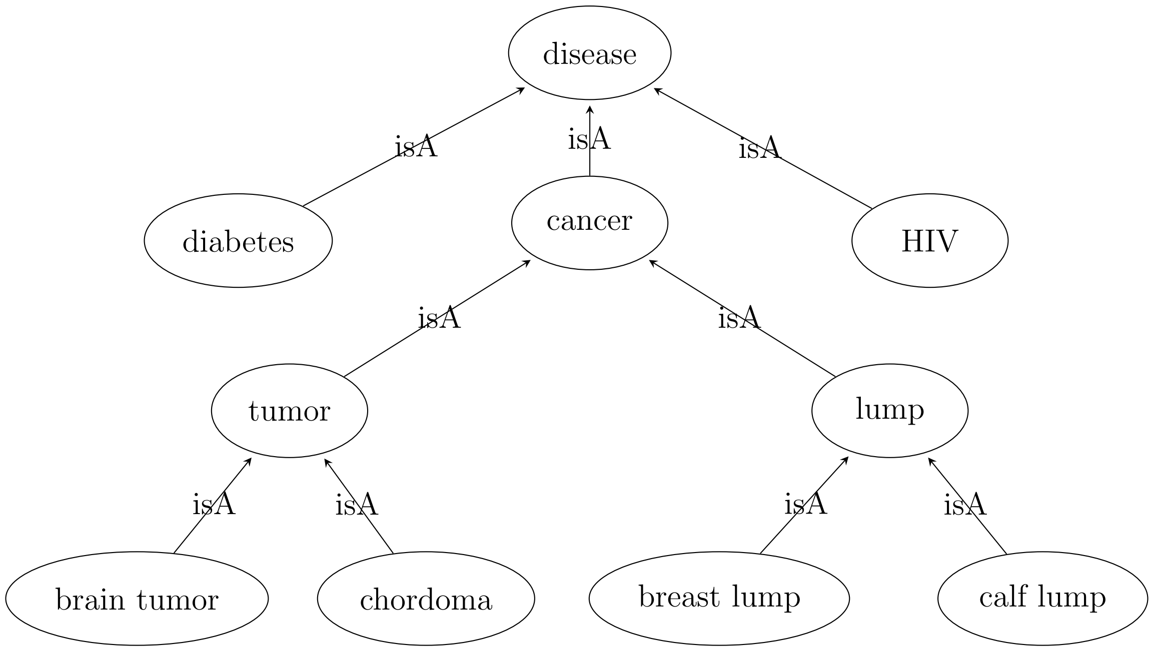

Fig. 5.3: Effect of Iteration on Number of patterns

more entities with a defined relation. Some relations are special in the sense that these relations express taxonomical information or INSTANCEOF information. These relations are specially useful for common-sense inference by means of special properties such as symmetry, functionality, injection, transitivity, reflexivity etc. We limited our work to exploring concepts and hierarchical relations in this chapter. As a future work, exploring properties of relations would add more knowledge and meaning to the bigger vision of concept based knowledge representation and extraction. We have made a simple approach of representing concepts and relations with lexical words or word-phrases as this is most close and intuitively understood by human. We are well aware that this leads to ambiguity of concepts because of synonymy and polysemy. As a future work, we would need to work towards the formalism of unique identifiers and name spaces as in semantic web.

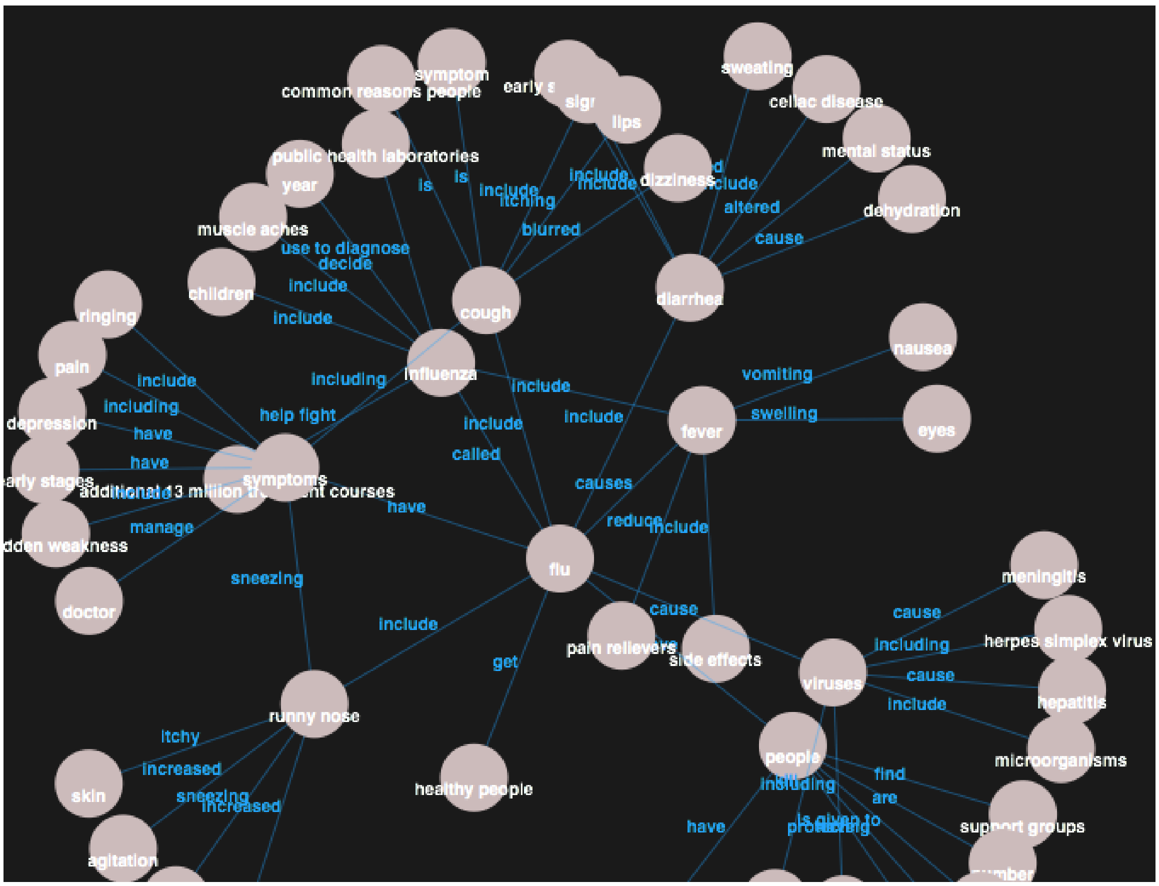

Fig. 5.4: A general view of Navigation-based Concept Retrieval in CREATE prototype

Chapter 6

CONCEPT BASED OPINION EXTRACTION

"We are so vain that we even care for the opinion of those we don't care for". –Marie von Ebner-Eschenbach

6.1 Overview

The usability and reliability of reviews have embarked them as a standard to assess the quality of products and services. Various studies have shown that online reviews have real economic values on the products that the reviews target (Dhar, Chang, & Stern, 2007; Ghose & Ipeirotis, 2011). The sheer volume of reviews spread over many sites makes it almost impossible to go through every one of them manually. This authenticates the need to develop an automated and intelligent tool that can process the reviews and turn them into useful information for users so that users can obtain the most useful information in least amount of time. Some of the review sites such as amazon.com have tried to solve this problem by presenting an average rating of the product based on the numerical value given by the users to indicate the sentimental orientation towards the product and showing *"The most helpful favorable or critical review.* This approach does not solve the problem completely as readers tend to limit the choice by only going through few reviews at the top. Moreover, different customers may be interested in different features of the product. For example, a frequently traveling person may want a light and small camera whereas a person who is a professional photographer may want to check resolution and other features. Thus it is hard for the reader to grasp overall picture of peoples sentiments. One of the major challenges that needs to be tackled to realize an automated review mining system is opinion detection and classification (e.g. positive, or negative). For example, *"I like this camera"* is a positive opinion and *"this camera did not produce good picture"* is a negative opinion. This is a much studied problem (Bansal, Cardie, & Lee, 2008; Eguchi & Lavrenko, 2006;

Esuli & Sebastiani, 2005; Pang, Lee, & Vaithyanathan, 2002; Pang & Lee, 2004; Turney, 2002). These works mostly focus on detecting the overall opinion of user towards the product. However, instead of detecting overall sentiment, it would be more beneficial to detect the opinion at a more granular level since a single review could have disparate opinions on different features of a product. For example, the sentence *"Although the picture quality of the camera is good, it is too bulky to carry"*, expresses a positive opinion on the *picture quality* feature of the product and a negative opinion about the *size* feature of the product. One could argue that ratable features are already mentioned in some reviews. However, most reviews tend to be free-form text and people tend to express opinions on different features even in a single sentence like in the above sentence.

Secondly, users tend to express their opinions about different features in a variety of ways. For example, the opinion sentences *"I loved their sushi"* and *"the bread was really fresh"* are talking about the targets sushi and bread. Although these sentences are not expressing opinions on exactly the same subject or target, they seem to be positive about the feature food of some restaurant. The example above illustrates two grave issues for an automated system; one is to identify target of sentence and the other is to identify what ratable feature, the target is referring to. The idea of using topic models and other dimension reduction algorithms such as LSA (Dumais, Furnas, Landauer, Deerwester, & Harshman, 1988), PLSA (Hofmann, 1999) and LDA (Blei, Ng, & Jordan, 2003) appear to be most suitable in unsupervised aggregation of semantically similar features in reviews and have been used in recent research work. However, these models are not precise enough in their naive form for practical usage and also do not cover most of the sentences as they are completely uninformed of the domain, features and lexical or syntactic information at the sentence or document level. Some form of heuristics has been introduced with sentence level LDA and choosing only nouns at the sentence level for better prediction (Brody & Elhadad, 2010; Titov & McDonald, 2008). However, still a lot of textual content that are not relevant to ratable features reside in sentences of reviews which introduce impurities in prediction. Another thread of research has exploited lexical information and co-occurrence information at sentence level. Some of these approaches are association rule based filtering (Hu & Liu, 2004) and double propagation algorithm (Qiu, Liu, Bu, & Chen, 2011). With the lexical or syntactic information, product or service targets and features can be extracted with very high precision and recall. However, these methods can only extract features that are explicitly identified by certain words. As an example, from our analysis in camera dataset, only 36% of sentences have explicitly mentioned the features, which seem very normal and usual in human language (Qiu et al., 2011). Another important part is that people by nature use different words or phrases to refer to single feature, which makes it virtually impossible to only use this method to extract most prominent features from reviews. As the number of reviews increases, the numbers of targets increase exponentially too. As an example, in the restaurant dataset, we extracted 10633 numbers of targets from 275512 numbers of sentences in 52624 reviews. Usually, reviews are in large number for most of the products and services and an effective review mining system should be able to provide the collective impression in shortest possible way. Thus, the real use of these extracted targets is only seen when they are grouped to few important features. In our approach, we propose a unified model that incorporates target and feature detection together at sentence level. With such a model, we can also tackle the problem of implicit and explicit mention of features in the review. To explain this, let us analyze the following sentence, *"I've had a lot of cameras, but this one is ACTUALLY being used, and not stored in a desk: it's always in my shirt pocket"*. The target of this sentence is possibly usage whereas the sentence is actually talking about the feature size. Thus, if we try to extract the target first and then try to relate that target to ratable features separately, the feature size can never be inferred from the sentence. Both the topic model based approaches and lexical or syntactic information based approaches, as explained above, have their own pros and cons. However, no work has been done to combine both the information to get the best of both. Hence, in this work, we tackle the problem exploiting both the local and global features preserving domain independency and least supervision. We came up with a strategy to filter the words in SLDA(Sentence LDA) so only relevant words are left behind. These filtered sentences are then fed to LDA to get co-occurrence based grouping of features. The only human intervention in our framework is the assignment of meaningful names to LDA discovered topics. Our results show distinct advantage over both the standard methods combining both the advantages and mitigating the disadvantages. The application of our model is multifold. It becomes an inevitable part in feature based summarization of reviews. It can also be essential in extracting supporting textual cues for rating of products. Feature based sentiment analysis can also be more accurate and representative of existing data as it projects detailed information. Moreover, since sentiment and opinion are totally subjective and personal to individuals, it might be better to present overall impression rather than the polarity in terms of positivity and negativity as polarity does not provide details on why it is good or bad. A feature that is positive for a person might be negative for the other. Hence to better handle such issues, we first detect targets mentioned in each sentence of reviews and group them into major ratable features for overall impression.

In earlier chapters, we reviewed and proposed methods to process factual data. While the methods work very well for factual data, it may not be the best method to handle subjective data. A single instance of highly confident factual statement is enough for factual retrieval whereas, in the case of subjective information retrieval, the desired knowledge is almost always, an estimation of collective wisdom for the concept or an aspect of a concept. Hence, we propose to deal with subjective data in a different way. This area of research where we retrieve opinion oriented (subjective) information is popularly known as *opinion mining and sentiment analysis*.

6.2 Problem Statement

We use the term *entity* to represent the products, services etc., that is being reviewed. An entity $|E|$ can be represented with countable set of ratable features $|F|$. Each feature f has a finite set of words or phrases associated with it. These words represent the feature of same class and are either synonymous, homonymous or hyponymous to each other.

An opinion holder comments on a subset of the features of an entity in a review r. Each review r has a set of sentences $|S|$ and a numerical value n indicating the overall sentiment of the opinion holder. Each sentence or sub-sentence s contains a target t (main subject of the sentence) in reference to one of the features f of the entity and sentiment polarity which is positive, negative or neutral on the target Note that some modifier words such as not may reverse the polarity of the opinion and this needs to be handled. Notice that the same word may have different polarities for different entities. It is even possible that the same word denotes different polarities for the same entity or feature. For instance, some people like big cars, but some find it a nuisance.

Here, a sentence or a sub-sentence can be fully represented in terms of three values (opinion-holder; target; polarity). To simplify the problem, we ignore the opinion-holder and assume that in all the sentences, the reviewer writes his or her own views on a product. This assumption is valid in the case of standard reviews as most people write their own experience. Thus, for each sentence or sub-sentence in a review, we need to identify target of the sentence in reference to one of the features and the polarity of the sentiment expressed (positive/negative). After extracting these tuples from each sentence, we need to group like targets together to classify them to one of the ratable features of the entity such that $\{t_1, t_2.t_n\} \in f$.

Finally, the system needs to be able to classify unseen sentences into corresponding features without going through target and sentiment identification step.

6.3 Related Work

The problem of opinion oriented information extraction can be seen as a combination of tasks such as subjectivity identification, sentiment detection, aspect identification and summarization. Based on the objective of the application, one or more of these tasks need to be performed to build a typical opinion mining related system. We will review some notable works in each of these tasks in the following section.

6.3.1 Subjectivity Identification

Subjectivity in our context is defined as a process of identifying whether a text of interest is expressing opinions and evaluations on a concept or concepts (J. M. Wiebe, Bruce, & O'Hara, 1999). Some documents, such as product reviews may be explicitly subjective as people rarely write any factual sentences in reviews. However, other documents such blog articles, news pages and internet forums may contain a mixture of factual and subjective sentences. The task in this case is to identify if the text is expressing some opinions or stating a fact. This discipline and sentiment detection task has been found to be closely tied together (Bruce & Wiebe, 1999; J. M. Wiebe et al., 1999; J. Wiebe, Bruce, Bell, Martin, & Wilson, 2001; Pang & Lee, 2004; J. M. Wiebe, 2000; Yu & Hatzivassiloglou, 2003). Subjectivity classification can prevent the polarity classifier from considering irrelevant or even potentially misleading text. Pang and her colleagues found that subjectivity detection can compress reviews into much shorter extracts that still retain polarity information at a level comparable to that of the full review (Pang & Lee, 2004).

Much of the research in automated opinion detection has been performed and proposed for discriminating between subjective and objective text at document and sentence levels (Bruce & Wiebe, 1999; J. M. Wiebe et al., 1999; J. M. Wiebe, 2000; Pang & Lee, 2004; J. M. Wiebe, 2000; Yu & Hatzivassiloglou, 2003).

6.3.2 SentimentEZ , Polarity Detection and degrees of positivity

Sentiment polarity detection is one of the core tasks in the process of concept based opinion extraction. In this discipline, given a text of interest, the task is to identify whether the text supports or opposes the target concept. Traditional classification methodology cannot be easily applied for this task as simple keyword based methods is not directly applicable and thus it may require technologies from natural language processing, information retrieval, information extraction etc. The work on sentiment detection started from the 1990s and several international conferences such as ACL, AAAI, WWW, EMNLP, CIKM have devoted special issues to this topic (Argamon, Koppel, & Avneri, 1998; Kessler, Numberg, & Schütze, 1997; Spertus, 1997). However, it started getting significant attention in the early 2000s (Dimitrova, Finn, Kushmerick, & Smyth, 2002; Durbin, Richter, & Warner, 2003; Chaovalit & Zhou, 2005; Gamon, Aue, Corston-Oliver, & Ringger, 2005; Glance, Hurst, & Tomokiyo, 2004; Grefenstette, Qu, Shanahan, & Evans, 2004; Hillard, Ostendorf, & Shriberg, 2003; Inkpen, Feiguina, & Hirst, 2006; Kobayashi, Iida, Inui, & Matsumoto, 2006; Liu, Lieberman, & Selker, 2003; Rauber & Müller-Kögler, 2001; Riloff et al., 2003; Subasic & Huettner, 2001; Vegnaduzzo, 2004; J. Wiebe & Riloff, 2005; Wilson, Wiebe, & Hoffmann, 2009). Talking about the method employed for the classification task, initially machine learning and semantic analysis techniques were used. Later, for better performance, natural language processing features such as POS tagging, dependency parsing were also used. Initial work on sentiment detection assumed the problem to be a binary classification problem where the text was assumed to be either positive or negative.

However, to make a more refined analysis, the problem can also be seen as multiclass classification problem that distinguishes degrees of positivity and negativity (strongly positive, medium positive etc). Koppel et al., stressed the importance of introducing a class for neutral text (Koppel & Schler, 2005). They showed that learning from negative and positive examples alone will not permit accurate classification of neutral examples. Moreover, the use of neutral training examples in learning facilitates better distinction between positive and negative examples. From the text size viewpoint, some works tried to detect the polarity of the whole document, which is generally talking about a single concept. Another thread, on the other hand, performed more fine grained detection, classifying each sentence of the document.

6.3.3 Opinion summarization

The sheer volume of reviews on products and services spread over many sites makes it almost impossible for users to go through every review manually. As a solution to this problem, opinion summarization systems were proposed (Ku, Liang, & Chen, 2006). These systems summarize the opinions of articles by detecting sentiment polarities, identifying aspects (attributes of the product) and generating a gist of all. This helps customers to get a collective impression on the product or service or event or topic and also helps the manufacturers understand how customers feel about it. One of the early works in the discipline was done by Hu et al. where, given a set of customer reviews of a particular product, they first identified the features of the product, identified the sentences that talked about the feature and detected their polarities and then produced a summary using representative sentences (Hu & Liu, 2004). Another group, investigated news and blog articles. For the purpose of experiments, they chose articles that focussed on animal cloning and proposed methods that could handle extraction at word, sentence and document level (Ku et al., 2006). They also take into consideration, the task of subjectivity identification. They then generated summarizations with representative sentences. Another work done proposes to generate rated aspect summary of short comments based on the overall rating provided (Lu, Zhai, & Sundaresan, 2009). They propose to use unstructured and structured PLSA to identify group of words representing aspects.

6.3.4 Applications

The applications of the methods we talk in this chapter range from the automatic extraction of attitudes and sentiments from a variety of texts such as government, political, entertainment and commercial product data in blogs, forums and reviews. It is obviously a significant part of our concept based information extraction. A significant amount of queries are seeking opinions related to specific products. Existing search based systems treat all the documents in the same way and retrieve most relevant pages according to the query. Sites such as epinions.com and Google products etc, attempt to make this process easier by aggregating reviews in a single place. This still does not solve the problem completely as there still is humongous number of reviews for user to go through. Similarly, blogs are also well known to be rich in opinions. In this chapter we will review some notable works in the field of opinion mining in reviews and blogs.

1. **Opinion analysis on reviews**

   Product reviews are one of the most explicitly opinion rich documents. Due to this speciality, the task of subjectivity detection can be validly ignored in reviews. Opinion analysis on product or service reviews was one of the earliest works done in the area of opinion mining.

2. **Opinion analysis on other subjective documents**

   Opinion analysis has also gained attraction in analyzing other subjective documents that revolve around a particular abstract concept such as politics etc (Adamic & Glance, 2005; Mullen & Malouf, 2006; Yan et al., 2009). Web logs contains a vast amount of information that could be used for different purposes such market intelligence, government intelligence, financial analysis etc. However, the task of opinion analysis in blogs is not as straightforward as in product reviews.

6.4 Proposed Methodology

The ideal goal here is to understand the semantics and sentiment of each text and then generate opinions summarized in terms of whats positive and whats negative. The problem when understood in an intuitive way requires collection of articles related to the subject, extraction of sentences expressing the sentiment about the subject, identifying the polarity of sentence, and then aggregating the total sentiment in terms of negativity and positivity. This preferably involves techniques of multi-document summarization, anaphora resolution, sentiment phrase identification, polarity classification, relevance ranking, etc. However, since we are dealing with a very large text, typically thousands of articles in a day, deep semantic analysis of each piece of text will be computationally very expensive.

The overall methodology can be seen as a combination of serial tasks such as domain specific seed sentiment words extraction, explicit feature candidate extraction based on sentence level features, assembly of target words to ratable features and classification of sentences to ratable features. In this section, we will explain each of the tasks in detail.

Figure 6.1 depicts the overall framework and flow of our system. Customer reviews on a product are collected from review sites in the web. Most of the reviews have ratings (a number indicating overall sentiment of the review) with them. We use this number to extract seed sentiment words in that domain. In the process, we do not consider the reviews that do not have ratings associated as this step explicitly needs a rating value. These seed words are then fed in the target extraction module which essentially extracts candidate target words and more sentiment words in a

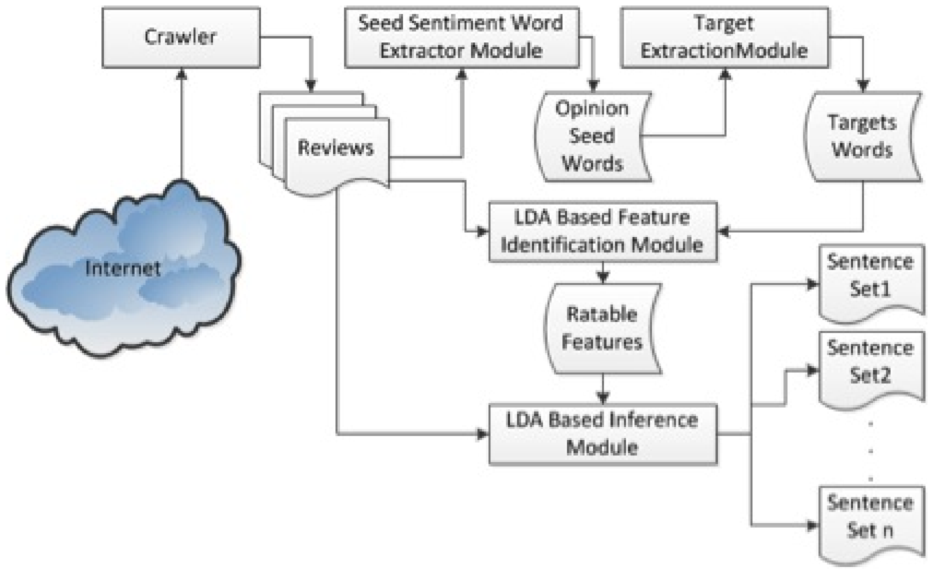

Fig. 6.1: An Overall System Framework for Concept based Opinion Extraction

recursion. These candidate feature words are then grouped by LDA based feature identification module to identify ratable features of product or service. We then use LDA based inference module to identify ratable features in unseen sentences.

6.4.1 Domain specific seed sentiment words extraction

While the original double propagation method requires an opinion seed set to start with, our method does not require any external information. Since opinion words are completely domain dependent, it might not be so easy to obtain domain dependent opinion words for hundreds of products and services. We exploit user given ratings to collect domain dependent subjective opinions. Sentiment is expressed most of the time with adjectives. Thus, we only work with adjectives in a sentence. Now for each filtered word in a review, we count its frequency in reviews which have positive rating and negative rating. A review is considered positive if it has a 4 or 5 star and is considered negative if it has rating of 2 or fewer stars. Here, we use majority voting to detect the underlying polarity of these sentiment words. We define the semantic orientation of a sentiment word based on its minimum frequency in positive or negative reviews and the ratio of its occurrence in them.

The semantic orientation is defined by the following equation.

$$SO(w) = \frac{|w_p|}{(|w_p| + |w_n|)} \tag{6.1}$$

Here, w_p is the frequency of word w in positive reviews and w_n is the frequency of it in negative reviews. If the semantic orientation of a word w (i.e. SO (w)) is greater than a given confidence in positive case and it appears at least in a given support value of positive documents, the word is considered to have a positive semantic orientation. Similarly, we also identify negative words. We changed the threshold values to obtain best precision and recall and hence found 60% and 2% as the optimal value. We also experimented in adverb and adjective phrase chunks with the same idea, and results showed that unigrams give best results.

6.4.2 Negation Handling

When considering a sentence as a bag of words, we lose the inherent structure of it and hence the semantic meaning of the sentence. A sentence may have a positive sentiment word; however it may have been modified by some negation word. For example, in a sentence, *"The food was not at all good"*, the sentence talks about the aspect food and has a sentiment word "good". However, the overall meaning gets reversed because of the word "not" in it. To solve this problem, we define a window size of 5. If a negation word appears within a window size of 5 and if the sentiment word is the first to appear after the negation word, we consider the opinion to be reversed. We use a predefined set of negation words to detect the presence in sentences.

6.4.3 Explicit feature candidate extraction based on sentence level features

The target extraction methodology is based on the syntactic relation between opinion words and the features. The process recursively discovers product or service targets based on known opinion words and discovers more unknown opinion words based on known targets.

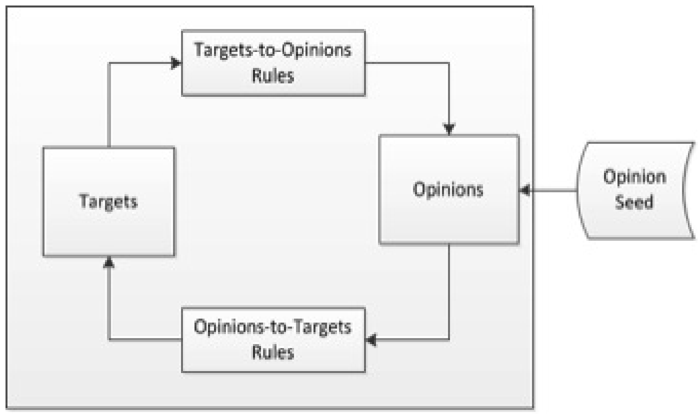

Fig. 6.2: Conceptual framework for Target Extraction Module

Figure 6.6 shows the conceptual framework of target extraction module. Initially, it requires a seed opinion word set, the extraction method of which has been described in the previous section. The process employs rule based strategy in dependency tree of a sentence. For example, if we know great is an opinion word and are given a rule like *"a noun on which an opinion word directly depends through mod is taken as the target,"* we can easily extract pictures as the target Similarly, if we know picture is a target, we could extract adjective *great* as an opinion word using a similar rule. The paper describes the extraction rules in detail (Qiu et al., 2011).

6.4.4 Assembly of target words to ratable features

Opinions expressed in reviews are mostly associated for particular features of the entity. Each useful sentence should have a feature mentioned along with an associated opinion. Thus the next step is to identify representative features of an entity in terms of groups of words that represent the feature. Latent Dirichlet Allocation(LDA) and its modifications have recently been applied to uncover the latent topics which are not directly observable in a document (Brody & Elhadad, 2010; Titov & McDonald, 2008). We also followed the same idea of using bag of words in documents to identify the aspects in the reviews. LDA is a generative probabilistic model well suited for discrete distinct data such as text corpora. LDA can be seen as a three-level hierarchical Bayesian model which models each item of a collection in terms of finite mixture over latent set of aspects. Each aspect is then modeled as an infinite mixture of aspect probabilities. This allows it to capture significant intra-document statistical structure. Documents are represented as random mixtures over latent aspects, where each aspect is characterized by a distribution over words. Since we are interested in the fine grained features of the entity, we assume each sentence to be a single document (Brody & Elhadad, 2010). Thus, the output of the model is a distribution over inferred aspects for each sentence in the data. We skip the details of LDA since it is out of scope in this work. Intuitively, features are mostly presented in noun forms. Thus, as a preprocessing step, Brody et al, filtered noun form of words from the sentences in reviews and use them as candidate to generate hidden features of the entity (Brody & Elhadad, 2010). We extend this filtering of words to another level in our work. Instead of using all nouns in the sentence, we filter the words to only keep candidate target words obtained from previous step to feed to LDA. LDA groups a set of representative words into pre-defined number of aspects. The following diagrams and the corresponding mathematical formulation represent sentence level noun filtered LDA and sentence level target filtered LDA.

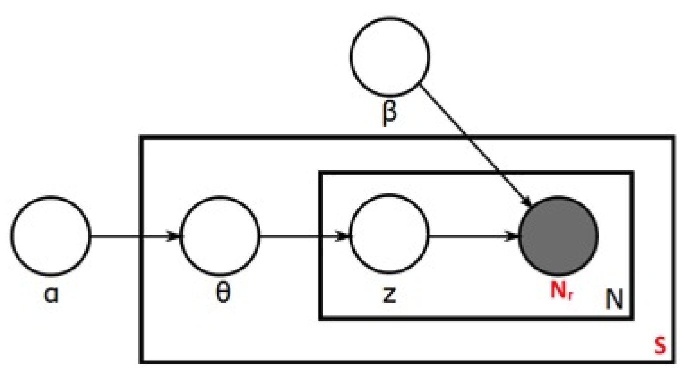

Fig. 6.3: Noun word based LDA

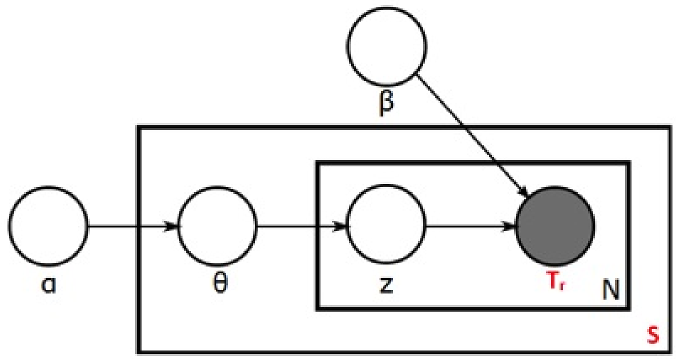

Fig. 6.4: Target word based LDA

We used the Gibbs sampling based LDA for estimating the parameters [21]. Let $\overrightarrow{N_r}$ and $\overrightarrow{z}$ be the vectors of all nouns and their topic assignment in the collection. Then, the topic assignment for a particular noun in the review is computed as follows:

$$p(z_i = k \mid \vec{z}_{\neg i}, \vec{N}_r) = \frac{n_{k,\neg i}^{(t)} + \beta_t}{\left[\sum_{v=1}^{V} n_k^{(v)} + \beta_v\right] - 1} \frac{n_{m,\neg i}^{(k)} + \alpha_k}{\left[\sum_{j=1}^{K} n_m^{(j)} + \alpha_j\right] - 1}$$

Similarly, let $\overrightarrow{T_r}$ and $\overrightarrow{z}$ be the vectors of all targets and their topic assignment in the collection. The topic assignment for a particular target in the review is computed as:

$$p(z_i = k \mid \vec{z}_{\neg i}, \vec{T}_r) = \frac{n_{k,\neg i}^{(t)} + \beta_t}{\left[\sum_{v=1}^{V} n_k^{(v)} + \beta_v\right] - 1} \frac{n_{m,\neg i}^{(k)} + \alpha_k}{\left[\sum_{j=1}^{K} n_m^{(j)} + \alpha_j\right] - 1}$$

where $\overrightarrow{\alpha}$ and $\overrightarrow{\beta}$ are the Dirichlet parameters, V is the vocabulary size $n_k^{(v)}$, is the number of times a topic k is assigned to the word v, $\neg i$ refers to the assignments excluding the current assignment.

Table 6.1: Phrase extraction rules and corresponding examples

Structure	**Example**
$ph \leftarrow \ni (NP, ADJP*)$	the vibe is hard
$ph \leftarrow \ni (NP, ADVP, ADJP)$	cakes are also excellent
$ph \leftarrow \ni (ADJP, NP)$	great fresh whipped cream
$ph \leftarrow \ni (NP)$ && $NP \leftarrow (JJ\ NN)$	wonderful dessert spells

6.4.5 Classification of sentences based on features

Each test sentence is given to the LDA model to get its topic distribution. The topic distribution is then used to classify the given sentence into one of the aspects. To this end, w we say that a sentence belongs to a feature ‘f’ if P(‘f’) >threshold, where P(‘f’) is the probability of topic ‘f’ in the sentence.

6.4.6 Representative Phrases Extraction

Each review can be seen as a collection of sentences. Each sentence can be represented in terms of aspect it mentions (if any), polarity of opinion expressed, and a meaningful phrase that gives a textual clue as evidence.

A **(representative phrase)** is a subset of a sentence, if it has the structure shown in the Table 6.1.

To extract representative phrases from sentences in reviews, we first chunk the sentence to obtain phrase chunks. Then we try to extract subset of phrase that matches the above conditions. These extracted phrases serve as candidates for textual cues presented in final summary.

6.4.7 Reviews Summarization

For a particular instance of an entity, we may have hundreds or thousands reviews which make it difficult to read. Thus our goal here is to generate a succinct statistical representation of likes and dislikes for different aspects of the entity along with supporting textual clues to have clear explanation.

To generate this statistics, we collect list of representative phrases for each aspect and for each sentiment. Then we calculate the similarity of two phrases in the list. If the sentiment words really mean the same thing, we choose only one phrase as the representative, and then increase the frequency of the phrase. We use WordNet to calculate the similarity between the words in two phrases. We only compare adjective with adjectives, and noun with nouns. At the end, we will get a few phrases with high frequency and few with low frequencies.

Table 6.2: Overall sentiment summarization process

Sentiment Summarization Algorithm
Input: instance class C, set of reviews R for C
Output: like and dislike votes for each aspect along with supporting textual clues
$\forall review\ r \in R$
$S \leftarrow extractSentences(r)$
for each sentence $s \in S$
$t \leftarrow extractTripleFromSentence(s)$
$T \leftarrow addToList(t)$
for each aspect
for each sentiment
$T \leftarrow collapseSimilarPhrases(T)$
Here $t \leftarrow \{(a, p, p_h)\} = f(s)$
where $a \in A(A \leftarrow setofAspects)$
$p \in P(true, false)$
$ph \leftarrow representativephrase$

The phrases with high frequency counts are representative ones that represent most peoples opinion. Thus, we can present these phrases with statistical support to show how many people are thinking that way about the entity. The following figure describes the process in detail.

6.5 Experimental Results

In this section, we will present our result in camera and restaurant domain for different tasks.

6.5.1 Dataset

To demonstrate the proof of concept of our model and its domain independence, we chose two completely disparate domains, restaurant (service) and camera (product). For the restaurant dataset, we used the existing dataset from (Ganu, Elhadad, & Marian, 2009). It consists of 652 reviews of restaurants in New York City, of which a subset is annotated with features and orientation, which we use extensively during evaluation. For the camera dataset, we collected reviews from amazon.com to develop the framework. For the evaluation purpose, we used data from Bing et al. (Qiu et al., 2011). We manually annotated the sentences in the dataset to 9 ratable features. The dataset was annotated by 2 graduate students of computer science. Some useful statistics of the dataset is shown in the following table.

For the preprocessing, we used Porter Stemmer to stem words in the text. We used OpenNLP toolkit to perform part-of-speech tagging of the text and generate phrase chunks from sentences (Baldridge et al., 2004). Some useful statistics of the dataset is shown in the Table 6.3.

6.5.2 Seed Sentiment Words Extraction

Table 6.4 below shows some sample sentiment words that were detected as positive and negative based only on provided user rating using all the reviews in

Table 6.3: Dataset Summary

Dataset	**Restaurant**	**Camera**
# of Reviews	52624	3999
# of positive reviews	38653	3212
# of negative reviews	9305	522
Avg # of sentences in positive reviews	5.09	4.93
Avg # of sentences in negative reviews	5.78	6.10
Average rating	3.946	4.17

both the dataset

Table 6.4: Sample sentiment words from constructed sentiment dictionary

Polarity	Sample Words
Positive	absolute, luscious, golden, refreshingly, cozy, amazing, tastefully, phenomenal
Negative	unidentifiable, tasteless, unfriendly, unprepared, unaccommodating, discourteous, salty

We evaluate our method using the manually annotated dataset provided by (Ganu et al., 2009). Each sentence in the dataset is positive, negative, neutral or has conflict. Among 3000, 2603 of them sentences are either positive or negative and we only use these for evaluation.

6.5.3 Candidate Target Extraction Based on Double Propagation

Since we did not have any conceptual contribution on the double propagation algorithm, we did not explicitly evaluate this task. However, we did apply the algorithm in a bigger and more realistic size dataset In the following table, we

Table 6.5: Evaluation result on seed sentiment word extraction

Polarity	Accuracy	Precision	Recall
Positive	76.46	83.74	89.04
Negative	73.69	100	71.48

present the statistics on number of targets generated by the algorithm on camera and restaurant domain. We also show sample target words extracted in both domains in Table 6.6.

Table 6.6: Target extraction statistics and sample target words

Domain	Sample Candidate Target Words	Targets	Sentences
Restaurant	Medallions , dining, roco, halibut, rock, steamy, souffles, omakase, tolerance, addict, border, caipirinas, sterling, vibes, vingear, vegeterian, shrimps, rigatoncini, soufflee, testosterone, mens, tostades,volleyball, ment, menu, job, woodsy, hamburger, shoulders, scent, passageway, sandwiches, verification, brusque, scene	10633	275512
Camera	Plastic, cravings, incentive, zoom, competitors, desktop, parts, video, purchases, amateur, photo, chance, party, viola, pocket, aspect, touch, washed, frames, nephew opening, quality	1459	20012

6.5.4 Grouping of Targets to Ratable Features

Tables 6.7 and 6.8 show some sample words that were identified by the LDA model.

Table 6.7: Sample words representing ratable features in *restaurant* domain

Topics	**Noun Filtering Based Grouping**	**Target Filtering Based Grouping**
Price	Food, service, prices, price, quality, ambience, atmosphere, bit, ok, nothing, size, great	worth, prices, bit, price, quality, portions, service, cheap, reasonable, average, pricey,value
Ambience	Food, service, staff, atmosphere, Service, wait, dcor, waiters, Very, attentive, bit, waitstaff	Bar, room, dining, music, dcor, tables, area, space, atmosphere, cozy, nice, seating
Food	Menu, sushi, dishes, Food, fish, chef, everything, variety, dish, items, specials, choices, tasting	Pizza, steak, cheese, meat, side, burger, taste, fries, plate, chicken, burgers, bread, sauce
	Wine, drinks, meal, list, selection, appetizers, bar, glass, drink, course, bill, entrees, bottle, beer	Menu, wine, dishes, list, appetizers, selection, everything, entrees, glass, course, bottle, tasting
Anecdote	Dinner, night, time, friend, lunch, friends, day, brunch, birthday, party, evening, boyfriend, group	Place, recommend, love, fun, date, nice, friends, perfect, people, eat, enjoy, spot, big, anyone
Misc	Restaurant, experience, times, time, reviews, dining, years, place, meal, visit, couple, NYC	restaurants, experience, places, favorite, city, top, dining, years, neighborhood
Service	Table, waiter, minutes, time, order, people, reservation, hour, waitress, manager, hostess, wait	Table, wait, minutes, order, reservation, bar, waiter, hour, waiting, reservations

For the evaluation purpose, we used the manually annotated dataset by (Ganu et al., 2009). They have annotated around 3000 sentences from 652 restaurants for its sentiment polarity and aspect. For each sentence, they assign one or more aspects among six aspects; food, staff, ambience, anecdotes, miscellaneous and price. For the evaluation on camera domain, we used Bing data which we annotated for 9 features and we used amazon.com collected data for target extraction and grouping.

Table 6.8: Sample words representing ratable features in *Samera* domain

Topics	**Noun Filtering Based Grouping**	**Target Filtering Based Grouping**
Shooting Mode	Mode, auto, settings, focus, feature, photos, setting, iso, results, modes, control, not, thing	Shots, light, flash, shot, shooting, speed, shutter, night, indoor, lighting, grainy
Video/ memory	Video, hd, flip, videos, camcorder, hours, anything, minutes, sound, ultra, movies	Video, videos, software, camcorder, download, record, sound, recording
Build/look	Camera, reviews, lot, model, people, Kodak, things, review, research, others, problems	Settings, feature, auto, focus, photos, button, setting, macro, automatic, function
Misc	Camera, reviews, lot, model, people, Kodak, things, review, research, others, problems	Camera, features, manual, user, learning, controls, functions, average, version
Anecdote	Pictures, camera, pics, kids, lot, work, days, course, fun, expectations, detail, kind	Pictures, family, daughter, sharp, vacation, friends, friend, crisp, show, pix, wedding
Size	Camera, size, use, pocket, something, everything, ease, purse, hand, weight	Camera, size, screen, compact, weight, bigger, portability, viewing, body
Price	Nikon, amazon, product, money, purchase, price, nothing, service, coolpix, deal, days	Good, price, deal, sale, reasons, buy, review, money, worth, purchase, return, pay, cost
Picture Quality	Quality, picture, image, pictures, photo, images, stabilization, lighting, situations, friend	Quality, picture, image, photo, stabilization, stabilizer, crystal, piece, versatility, blast
Battery Life	Camera, battery, time, life, year, wife, daughter, Christmas, gift, great, son, problems	Battery, batteries, life, time, charge, portable, power, charger, stick, run, shirt

Figures 6.5, 6.6, 6.7, 6.8, 6.9, 6.12 above show the precision-recall curve for some of the ratable aspects in restaurant and camera domain. A threshold tf for each feature is defined to classify sentence to one of the inferred features. We assume a sentence as associated to feature f $if P(f)>tf$. By varying the threshold tf we created precision-recall curves. This is similar to (Blei et al., 2003). For a direct comparison to their method, we also implemented noun based filtering along

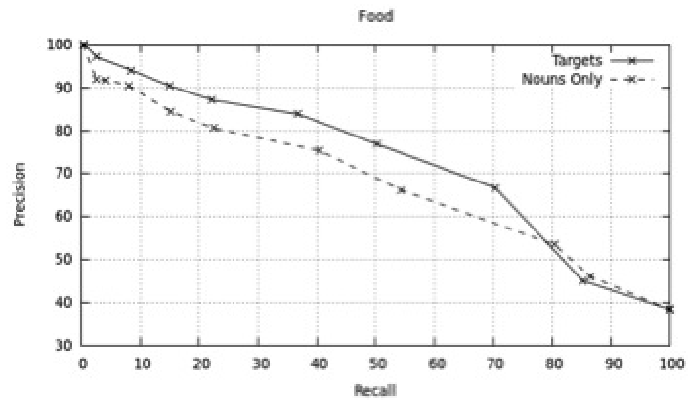

Fig. 6.5: Precision and Recall Curve for ratable aspect *Food* in restaurant domain with *noun and target filtering*

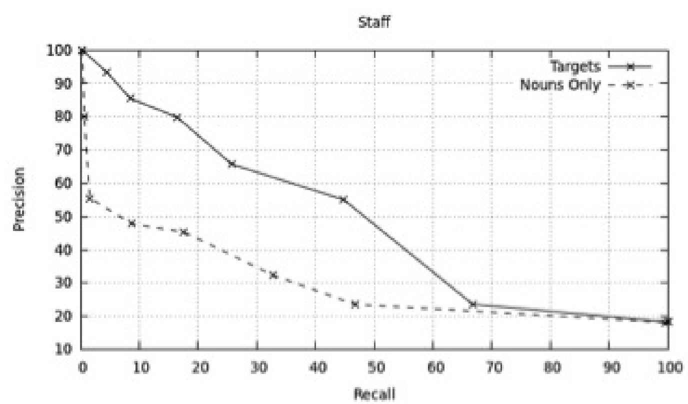

Fig. 6.6: Precision and Recall Curve for ratable aspect *Staff* in restaurant domain with *noun and target filtering*g

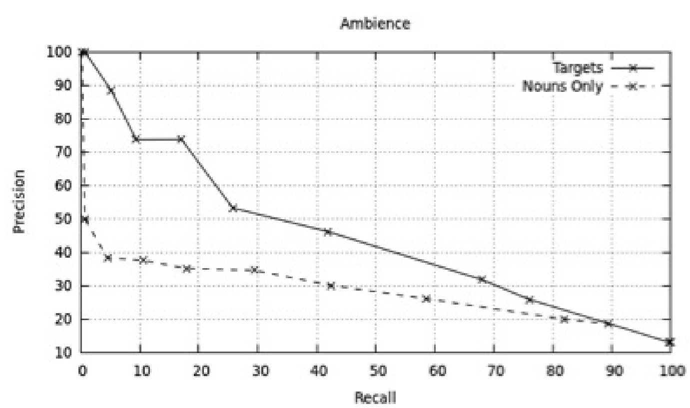

Fig. 6.7: Precision and Recall Curve for ratable aspect *Ambience* in restaurant domain with *noun and target filtering*

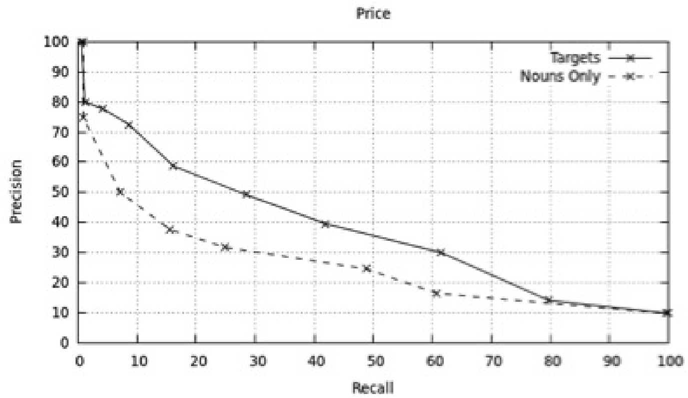

Fig. 6.8: Precision and Recall Curve for ratable aspect *Price* in restaurant domain with *noun and target filtering*

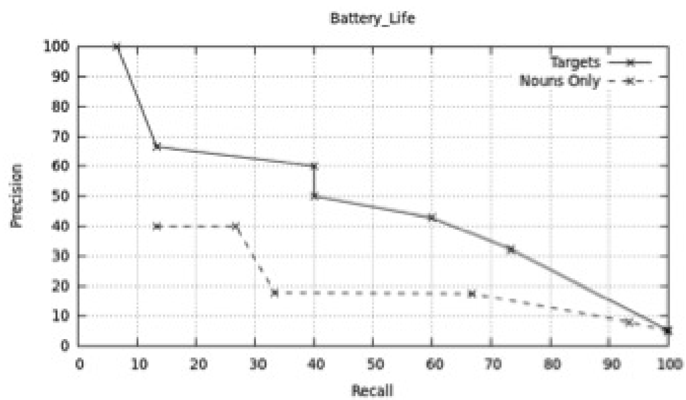

Fig. 6.9: Precision and Recall Curve for ratable aspect *Battery life* in Camera domain with *noun and target filtering*

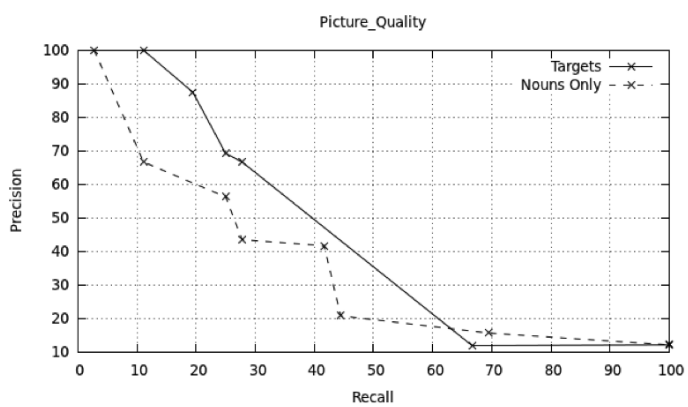

Fig. 6.10: Precision and Recall Curve for ratable aspect *Picture Quality* in Camera domain with *noun and target filtering*

with target based filtering and drew the curve that depicts significant improvement in precision-recall with the target based filtering method.

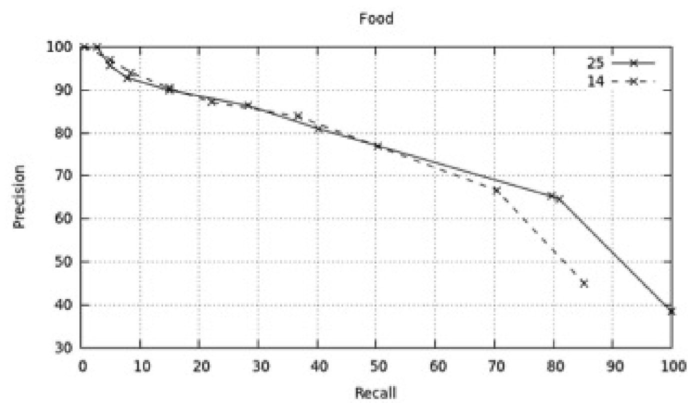

Fig. 6.11: Precision-Recall curves with 14 and 25 topics for ratable aspect *Food*

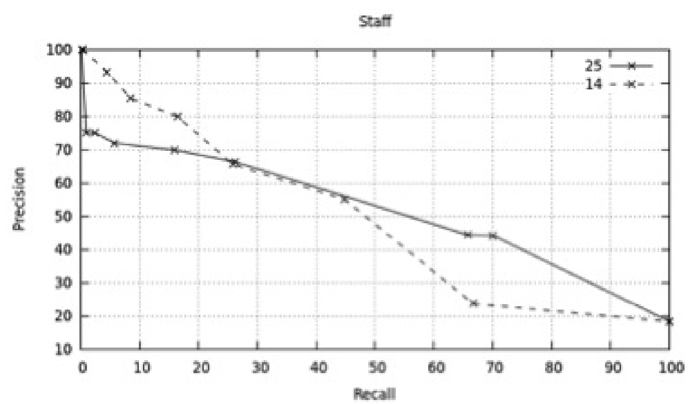

Fig. 6.12: Precision-Recall curves with 14 and 25 topics for ratable aspect *Staff*

The problem of determining the model order(number of features) still persists in our method as in most unsupervised learning scenario. We thus performed a small experiment varying the number of features in precision-recall graph as shown in Figure 6.12. As the result shows, although choosing right number of features(14 in case of restaurant domain) did seem to improve the result, the improvement was not significant. Hence, our systems is based on a practical number of features that users would think of, while reviewing a product or feature. For the restaurant domain to be able to evaluate our system we chose to use 14

Table 6.9: Sample Summarization Result

Aspect	Polarity	Summarized Cues
Food	Positive (222)	Fresh cupcake (48), sweet icing(12), right amount(11), rich texture(6)
	Negative (41)	Dry cupcake(11), cold glass(7), stale tasting(6)
Service	Positive (132)	workers nice(10), new people (9)
	Negative (23)	big disappointment(9), pretentious waiter(5)
Miscellaneous	Positive (129)	Good place(25), thick layer(12), near square(10), best bakery(9)
	Negative (31)	long lines(19), extraordinary demand(4)

features with the mapping of 6 manually annotated features. For the camera domain, we tried to manually annotate the dataset from Titov et al and found that there are nine practical features (Titov & McDonald, 2008). So we used nine features for camera dataset.

6.5.5 Review Summarization

We finally generated summarized textual clues with statistical voting to each clue. A portion of the generation is shown in Table 6.9.

6.6 Conclusion

We presented a framework to identify and aggregate ratable features with minimal supervision. Our method has shown significant improvement on the identification and grouping of features in reviews. We also introduce the idea of using the star rating as a way to classify sentiments without an external corpus. Since all the steps were performed in a domain-independent way, the system is flexible enough to be equally applicable to any other entity of any domain. Though, the recall of aspect identification system is not high, in real life scenario, most of the products or services have sizable amount of reviews and even a low recall result could be representative and helpful to customers. Our work is not limited to product or service reviews mining. The same infrastructure can be used to process opinion information for any possible concept from heterogeneous sources such as reviews and blogs, whichever is suitable. The proposed system, if built successfully, will have a broad impact in the field of opinion mining, summarization and information extraction. We will explain our overall concept based opinion extraction in the Figure 6.13.

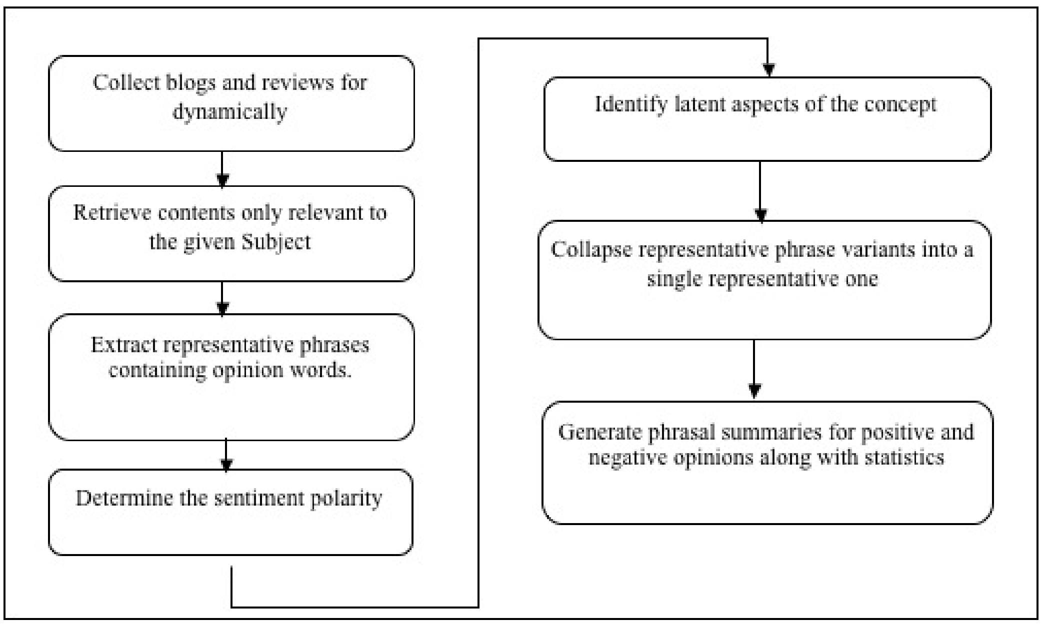

Fig. 6.13: Flow diagram of a baseline opinion extraction system

As future work, we intend to extend the system to collect all the reviews for a particular entity from the web and produce succinct information. This work can be seen as a milestone to build a system that does not perform keyword based document retrieval but actually processes relevant documents to produce precise information. Current work ignored the requirement of identifying opinion holder of the opinion. However, if the same concept needs to be applied in more general contexts such as blog to generate a summarized form of information for an entity such as a person or a company or location, opinion holder identification cannot be ignored. Moreover, in such a scenario, we would not even have explicit sentiment polarity clue such as rating. Thus an entity based faceted opinion summarization in such case would be more challenging. We intend to explore in that direction down the line.

Chapter 7

APPLICATIONS AND OPEN QUESTIONS

"I have not failed. I've just found 10,000 ways that won't work".– Thomas A. Edison

7.1 Overview

The capabilities provided by concept centric representation and extraction allows systems in natural language processing and information retrieval to offer better user experience with semantic representation of information. This technique is useful in a diverse set of applications. In this chapter, we will discuss about various potential applications along with some notable works done in the same area. Fast extraction of structured information has always been hard in the field of data mining and natural language processing as human language is capable of expressing same information in myriad of ways. Natural language is rich in synonyms, hypernyms and hyponyms which make the problem even harder. When thinking about the entities that could exist in the world, there is no count. Thus automated systems haven't yet been able to capture all the facts of the world about all the entities. We outline some of the prominent open questions that need to be answered to realize a complete machine reading system in the section following applications.

7.2 Applications

7.2.1 Information Retrieval

Most applications that are in use today are search oriented. The fact that more than seventy percent of search queries contain named entities suggests that people towards entities and their relations (Guo et al., 2009). For example, users typically want to see the product information using web searching engine. Named entity recognition is used to detect the intended entities in search queries (Guo et al., 2009). Once the entities of interest are identified, next step is to provide their details i.e. facts and relations. Recently, Google introduced *"knowledge graph"* which is used to show direct factual results for the entities of interest in their web search results (Singhal, 2012). It is used to improve search results using disambiguation of search queries, exploratory search suggestions and log-based summarization. Another example is Renlifang [1], a Chinese search engine for exploring people, locations, organizations and their relationships. Since IE provides facts about the entities, they are highly useful in information retrieval. Although open domain information provides the user, greater capability while retrieving information, it requires more effort to develop. Further areas of application for global relation extraction exist in fields like market analysis and in the creation of large knowledge.

7.2.2 Summarization

Summarization and information extraction share similar goals: both try to extract information from unstructured texts as per users' desire. The information in summarization, however, are natural language sentences unlike structured information in case of IE. Since most of the summarization tasks involve finding facts about entities, their relations and events, they are greatly benefited by information extracted by the IE systems. Typical use of entities and relations in summarization includes ranking sentences and removing redundancy in summary generation (Ji et al., 2013).

One of the early works that use IE in summarization is where extracted entities and events are integrated with natural language generation (Radev & McKeown, 1998). Similarly, Vanderwende et al. proposed the event-centric summarization by extracting portions of logical forms (Vanderwende, Banko, & Menezes, 2004). An unsupervised multi-document summarization technique for biographical summaries is proposed by Fadi et al. that exploits entities and time facts from Wikipedia (Biadsy, Hirschberg, Filatova, & InforSense, 2008). They

[1]http://renlifang.msra.cn/

report that sentence extraction is improved by extracted entities and facts. Hachey also justified use of generic relations to improve extractive summarization (Hachey, 2009). Recently a work, generated the abstractive summaries only using the open domain IE based template filling (Ji et al., 2013). They concluded that pure IE-based system lacks coverage, accuracy and inference in summarization. However, integrating information extracted by IE system to the state-of-art multi-document summarizer significantly improved its performance on both standard summarization metrics and human judgment.

7.2.3 Question Generation

The basic intuitive process of question generation is to first extract facts in some form of formal language and then transform them into interrogative form. Since the web contains wealth of information, it is an attractive resource for getting answers to many questions, ranging from simple (e.g. factual questions) to more complex ones. The knowledge base constructed using IE and web can used to provide answers to users questions.

Srihari and Li believe that the ideal test bed for IE is question answering (QA) because many questions in QA task seek entities as answer (Srihari & Li, 1999). For instance, *who* expects person, *when* expects time or date, *where* expects location, *how long* expects duration or length and so on. They report more than 80% of question of this type in the TREC-8 QA dataset Although they are not sufficient to answer a question, the gist of using named entities in their IE-supported QA is to narrow down potential text section containing answer in documents. The authors found significant improvement on QA by using IE on it. An open-domain question answering system such as True Knowledge heavily rely on the structured information extracted from the web (Tunstall-Pedoe, 2010). One potential challenge of using the web and IE in QA is to discard misinformation contained in the web.

7.2.4 Question Answering

On the verge of machine reading, question answering and IR have common interests. The underlying intention in information retrieval is to obtain required information in shortest click while question answering explicitly demands interaction in first click. Intelligent machines are expected to provide specific answers to specific questions by understanding, synthesizing and reasoning from structured data obtained in information extraction process. Ravichandran and Hovy introduced the idea of using patterns to match answers for a question based on relations and then used these patterns to answer more questions (Ravichandran & Hovy, 2002). Other professional efforts for semantic question answering are Project Halo[2] which creates scientific knowledge base that can answer questions and explain answers. The Semantic Research Assistant can answer ad hoc questions for clinical researchers based on semantic database (D Pierce et al., 2012). The main components are a content repository driven by a meta-model persisted in XML format, an inference based natural language query interface and data production model. Watson [3] is a computer question answering system developed by IBM to play and win the game of jeoparty against human. TextRunner, as discussed before, is another natural language question answering system that answers questions based on extracted tuples from web.

7.3 Open Questions

7.3.1 Knowledge Representation and Management

Web-scale open domain extraction poses itself with obvious challenges of tuple representation, organization and maintenance. In addition to the old problem of information retrieval such as document authenticity and priority, it now also needs to solve the problem of efficient representation of tuples in persistent layer

[2]http://www.projecthalo.com/

[3]http://researcher.ibm.com

due to added structure of relation tuples. Whether to organize the tuples as sequence of extracted tuples in document level for further processing, or to construct a hierarchical ontological structure at a hoslistic level is still an open question. Semantic units of information extracted by open domain information extraction process is likely to be highly noisy and imprecise. This illustrates the need for management of noisy data with the associaton of confidence values that represent the degree of correctness. However, it is not obvious on how to assign numerical confidence values with each extraction and how to choose an efficient imprecise data model for storing extraction results.

7.3.2 Entity Canonicalization

Natural languages have properties of synonymy, polysemy, hypernymy, hyponymy etc. Due to these properties, different concept identifiers in natural language may refer to the same real world entity or same lexical entity could mean different real world entities based on its context. In order to build a consistent and complete system, these properties of concepts and relations need to be resolved. We extract entities and relationships from unstructured source that needs to be integrated to existing database. The main challenge in this task is to decide if two strings refer to the same entity in spite of many noisy variants in which it appears in unstructured source. This problem is a hot area of research and is also known as *entity matching, record linkage, and de-duplication* (Shen, Li, & Doan, 2005; H. Zhao & Ram, 2008). The fundamental idea behind this process is to obtain attributional similarities between entities of interest. What determines a good degree of similarity is still a question. This problem has been well studied in structured data with well defined schema (Bhattacharya & Getoor, 2007; Singla & Domingos, 2006; Pasula, Marthi, Milch, Russell, & Shpitser, 2002). However, no work has been done with a focus in openly extracted noisy data. This problem has been formulated as classification problem in (Newcombe, Kennedy, Axford, &

James, 1959) and it has been noted that in addition to attribute similarities, the deduplication will be efficient if we also include information on what kind of relationship the entity participates in (Ananthakrishna, Chaudhuri, & Ganti, 2002).

7.3.3 Relation Canonicalization

Equivalent semantic relation can be expressed in many different forms. For example, in relation tuples (Google, is located in, Mountain View) and (Google, is based in, Mountain View), the semantics of both the relations *is located in* and *is based in* is identical although located and based may not be identical in many other cases. Relational similarity between two pairs of words is defined as the correspondence between semantic relations that exist between the two words in each word pair. Although this makes a huge difference in the improvement of information extraction, it is not widely performed due to the challenges. One of the prominent challenge to detect relational similarity is that entity pairs could have more than one relation between them. Moreover, entities themselves also need to be disambiguated based on the context.

Turney proposed Latent Relational Analysis(LRA) to address this problem by extending the vector space model with extracted patterns, SVD to smooth frequency of data and synonyms of words (Turney, 2006). Bollegala et al., proposes Relsim, an algorithm to calculate similarity between relations by first finding the context of the entities, then extracting patterns from these contexts and clustering them together to obtain the information for similarity (Bollegala, Matsuo, & Ishizuka, 2009).

7.3.4 Tuple Relevance

Relevance ideas developed in traditional information retrieval may not be appropriate for search in relation tuple based retrieval system. The challenge here is to construct an entity centric result. In the concept-centric paradigm, the interest is in the semantic relevance of relation tuples as opposed to traditional keyword based relevance. To evaluate the user satisfaction and select semantically relevant tuples is still a challenge to be addressed. The problem can be formalized as follows. Given an input query which could be a concept(subject or object) or relation and a large database of relation tuples, the problem is to decide, which tuple is most relevant among the total matches of tuples. Traditional retrieval method concentrate on the TFIDF similarity score, which has been found to be highly effective in text searches. The TFIDF similarity between two tuples t_1 *and* t_2 *such that* $t_i = (c_1, r, c_2)$ is given as follows:

$$TF - IDF(t_1, t_2) = \sum V(t_1.c_1, t_2.c_1)V(t_1.r * t_2.r)V(t_1.c_2 * t_2.c_2) \qquad (7.1)$$

$$V(q, c_1) = log(TF(q, c_1) + 1)log(IDF(c_1)) \qquad (7.2)$$

Concept centric relevance might be different than the vector space similarity as the most relevant entity for a query entity may be the one that co-occurs most with it. As an example, concept *"president"* might be the most relevant for the entity *"Obama"*. How to decide this relevance in case of semantic search is still a research question.

7.3.5 Tuple Filtering

Most of the existing works to date, have performed frequency (such as frequency of tuples, PMI of entity pairs) based accessing of tuples and specificity based accessing of patterns (in methods of pattern based extraction) (Yates et al., 2007; Fader et al., 2011). TEXTRUNNER uses a redundancy based method (Yates et al., 2007). The system first normalizes like tuples by converting the tuples to base form using morphological analyzers. After having tuples normalized, the system counts the number of distinct sentences from which each extraction was found. These counts serve as a measure of confidence that a tuple is a correct instance of a relation among entities. Extractions with a count of one are not added to the systems knowledge base. Reverb uses logistic regression classifier to assign a confidence score to each extraction, which uses multiple features which are efficiently computable and relation independent (Fader et al., 2011). The system is trained with a set of 1000 sentences from the web annotated by human annotators. These filtering methdologies discussed however, do not address the problem of tuple coherence and semantic validity. One of the obvious paths to explore is inference based filtering to select semantically sound tuples. As an example, the tuples *(X, is located in, Y)* and *(X, is mother of, X)* cannot exist for same X.

7.3.6 Commonsense Inference

Human mind has a special power to deduce some logical facts based on some known property. An automated system, on the other hand will only processes what it is given to. The system would thus need a component that would be able process some intuitive domain independent inference to make the system more powerful. This research area still has a lot to answer and remains an open challenge for the researchers. Inference is the process of extrapolating information that is otherwise not stated, based on stated information in the knowledge base. As an example, if the knowledge base contains tuples *(X, is located in, Y)*, *(Y, is located in, Z)*, the question is can systems infer the fact *(X, is located in, Z)* with fair generality to open domain concept. Schoenmackers et al. introduce the HOLMES system, which utilizes textual inference (TI) over extracted tuples by formulating a set of probabilistic inference rules as Markov logic Horn clauses, which can be used to derive new assertions (Schoenmackers, Etzioni, & Weld, 2008). Markov logic has been proposed to leverage joint inference and knowledge incorporation. This, however, comes with its own challenges (Richardson & Domingos, 2006). Theoretically, this is very effective and intuitive for machine reading system. Practically, however, it might lack on performance speed and quality. This is very important when working at web scale. Inference, in itself, is very complex due to the existence of sizeable number of hidden variables. The deep computational architecture projected by this system is thus hard to learn.

7.3.7 Higher Order Relation Extraction

All of the systems that we have been discussing focus primarily on binary relations. McDonald et al., have looked at complex relations and experimented for 4-ary relations (McDonald et al., 2005). The rely on a name-entity recognizer since the method assumes that entities and their types are known. Also, the order of the relation should also be known beforehand. Theoretically, TextRunner can also deal with n-ary relations based on n-ary training instances. However, it is not clear on how this system will perform and how to gracefully handle these complex relations still needs some work.

7.3.8 Temporal Aspects

Temporal knowledge in semantic world is an important parameter for systems such as question answering and information retrieval as it will offer more insight on when entities and their relationships were held. For example, *(X, is current president of, United States)* is a time dependent fact. In order to realize a complete system, we need to extend tuples to gracefully handle temporal aspect in it. There are several algorithms for classifying temporal relations between events (Mani, Verhagen, Wellner, Lee, & Pustejovsky, 2006; Chambers & Jurafsky, 2008; Yoshikawa, Riedel, Asahara, & Matsumoto, 2009). These are trained on TimeBank dataset (Pustejovsky et al., 2003). Timely YAGO proposes to extract temporal facts using regular expressions in Wikipedia infoboxes. This is very specific to Wikipedia infoboxes and is not applicable to arbitrary text (Wang, Zhu, Qu, Spaniol, & Weikum, 2010). Talukdar et al, proposed a graph based algorithm, that took into account the transitivity of temporal order within the document (Talukdar, Wijaya, & Mitchell, 2012). However, holistic ordering of tuples and validating tuples based on date still needs to be done.

7.3.9 Efficiency

Current web is highly volatile and highly dynamic as contents and expectations of users are changing all the time. Extraction models take time to be tuned towards the unstructured data. Thus, when the content changes, there is a learning curve for these systems. More web content is ever increasing and constantly changing, there is an obvious challenge to have a highly efficient extraction system that can meet the dynamicity of web. The extractions made on web need to be filtered to only select accurate tuples. The challenge, overall is to select the right document and hence, the sentence and then process it to extract tuples. The granularity of extraction, mainly at sentence level and preprocessing times add extra complexities to the system, and thus needs to be addressed.

7.4 Conclusion

In the current information age, the nature and usage of web is very dynamic and evolving. With increasing data being available, the value and importance of information is also ever increasing. Thus, it will be no exageration to state that systems that read text are highly required and demanded. In this chapter, we have shown, with major applications, the exploding potential of structured information representation. We have also tried to list out the most important research challenges that need to be solved before we can visualize a fully functional machine reading system. The growth in web data and the potential of this data both lead the research community to work towards the bigger vision of concept centric information extraction and management. This will be the next big evolution in information retrieval and machine reading paradigm. In this chapter, we shared our thoughts on the potentials and technical difficulties on the way to this evolution. This is a big opportunity for research community to make major contribution towards the emerging and foundational problems.

Chapter 8

CONCLUSION AND DISCUSSION

" This book fills a much-needed gap."– Moses Hadas.

8.1 Conclusion

The prominence of concept based knowledge representation is attributed to the unprecedented growth of heterogeneous data in the form of web-pages. Modern information technology is already greatly influenced by structured data counterparts. Search giants like Google and Bing have already started incorporating in their results, some form of objective results. An example snapshot of Google search results page for the query *"how to reduce stress"* is shown in Figure 8.1. Traditional keyword based information retrieval strategies assumed that users have some knowledge to formulate a suitable query that will retrieve required documents. However, with the increasing amount of web content and its usage by huge number of users, the search systems now need to be accessible and usable to any naive user. The user should not have to reformulate the query again and again to be able to retrieve the correct data. Having said that, we are also aware that language understanding still needs to solve complex parts and has to work together with knowledge representation and reasoning.

While this dissertation is mainly geared towards concept based information retrieval, there are many other applications within the scope of concept extraction and representation that can benefit from this fundamental proposition. Some examples are document classification, machine translation, data warehousing etc. It still needs to be noted, however, that concept centric retrieval and hence open domain information extraction is a very young, important and challenging problem in the field of NLP and IR. This is a fundamental problem that lives at the core of machine reading. This dissertation presents and discusses the core components in the life-cycle of concept based information processing in heterogeneous data. The

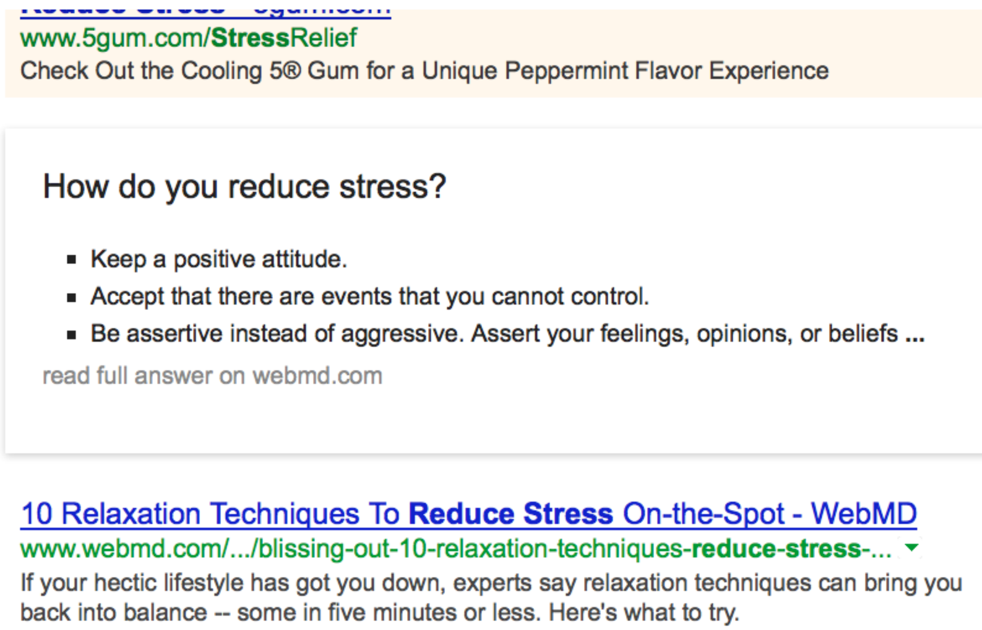

Fig. 8.1: Example of Google answering the question in Search Results Page

solution to the problem requires work on interdisciplinary fields such as machine learning, databases, web crawling, and information retrieval. The nature of the problem, where the data is domain agnostic suggests that supervised methods are not usable as a solution. We, thus introduced an unsupervised approach to the whole problem of concept based extraction and representation in factual as well as subjective data. The open domain information extraction methodology introduced in this dissertation was rule and pattern based. These methods are able to uncover structured information that natural language text is explicit about. To realize a successful concept-centric system and obtain rest of the information, many other challenges such as entity resolution and inferencing needs to be reliably solved. The process becomes more challening due to the existence of polysemous and synonymous concepts and relations. The work done in the dissertation is a step along the way towards broader vision of complete machine understanding task. The methodologies introduced are also fundamental to many IR and NLP problems such as question answering, summarization etc as discussed in previous chapters. In the earlier part of the dissertation, we focused on core models for the extraction of relational tuples via rule-based and induction based models. We first defined information extraction in general and open domain information extraction and then described different methods applied for open domain information extraction. We described rule based open domain information extraction along with its results and induction based open information extraction. To realize a complete concept based system, the first step is to extract and represent concepts. We defined concept as an entity corresponding to real life concept existing in human perception. As an example, person is a concept with its instances (entities) being Obama, Mark, Joey etc. The second step is to extract possible relations between these concepts that maps a concept to concept to form a connected multi-graph. The third step is then to organize these concepts in a hierarchical fashion to help in semantic information retrieval. With such an organization, the system will be able to make simple common-sense inferences. As an example, let us consider our system has a tuple , isBornIn(Martin Luther King, Atlanta) and isIn(Atlanta,Georgia). With these facts in the knowledge base, if someone asks isBornIn(who, Georgia), the system is able to retrieve "Martin Luther King" as a resultant entity. Similarly, we also propose to handle subjective (User Generated Content) data based on concept and its attributes. We describe the process on how main concepts can be discovered for a domain, how tuples can be organized in a hierarchical manner in chapter 5. Our main contribution is that we proposed an unsupervised way of extracting concepts and relations, organizing and retrieving them from knowledge base. Previous attempts have been made specifically in open information extraction and ontology based knowledge representation as separate tasks. In this dissertation, we try to amalgamate these two processes to realize a concept centric retrieval system. We propose to construct a framework not just for factual data, but also works for subjective data. We have qualitatively and quantitatively demonstrated the effectiveness and usefulness of the framework of concept representation and extraction. With increasing data being available, the value and importance of systems such as CREATE is ever increasing. Current surge of targeted marketing also corroborates imperative need to answer the open questions in information extraction by the research community. We have demonstrated the prospects of relation extraction systems. At the same time, we also need to be aware of the challenges that need to be solved before we can realize a fully functional machine reading system.

8.2 Future Work

Our goal in the dissertation has mainly been towards extracting explicit "worn on sleeve" semantic information from natural language text. In spite of a lot of efforts on designing models for extraction, no system has been able to attain impressive accuracy. Just be extracting explicit information, the recall is never going to increase. To realize a complete machine reading system, we should be able to infer implicit meanings hidden in a sentence or among multiple sentences. We haven't worked in detail on the inference possibilities that will complement extraction systems to uncover other hidden information in the text. There also needs to have some chain effect on the tuples extracted from the same sentence and hence same document. As a future work Markov logic network needs to be explored to study how joint inference model can be developed from ground tuples (Richardson & Domingos, 2006). More complex types of extractions problems such as entity resolution and word sense disambiguation, soft attributes of entities, and higher order structures, are only starting to be explored. As a part of future work, these fields need to be explored in more detail to build a complete system. Developing a machine reading system with current technologies need to deal with noisy and probabilistic data. In the simplest sense, we have seen that a confidence value is associated with each tuple to represent its correctness. However, we still need to deeply study on how these values interact with each other when it comes to dealing with multiple tuples either in hierarchical structure or in relational structure. This will also demand for the design of persistence systems that support managing, storing and querying uncertain tuples. There is a growing need to be able to flexibly represent the uncertainties in the data, and to efficiently query the data. Probabilistic databases have been of interest to solve such needs in recent times. We will need to study how this fits into the model of machine reading in the long run.

References

Adamic, L. A., & Glance, N. (2005). The political blogosphere and the 2004 us election: divided they blog. In *Proceedings of the 3rd international workshop on link discovery* (pp. 36–43).

Agichtein, E., & Gravano, L. (2000). Snowball: Extracting relations from large plain-text collections. In *Proceedings of the fifth acm conference on digital libraries* (pp. 85–94).

Ananthakrishna, R., Chaudhuri, S., & Ganti, V. (2002). Eliminating fuzzy duplicates in data warehouses. In *Proceedings of the 28th international conference on very large data bases* (pp. 586–597).

Argamon, S., Koppel, M., & Avneri, G. (1998). Routing documents according to style. In *First international workshop on innovative information systems* (pp. 85–92).

Atserias, J., Zaragoza, H., Ciaramita, M., & Attardi, G. (2008). Semantically annotated snapshot of the english wikipedia. *Proceedings of the Sixth International Language Resources and Evaluation (LREC'08)*.

Auer, S., Bizer, C., Kobilarov, G., Lehmann, J., Cyganiak, R., & Ives, Z. (2007). Dbpedia: A nucleus for a web of open data. In *The semantic web* (pp. 722–735). Springer.

Baeza-Yates, R., & Raghavan, P. (2010). Next generation web search. In *Search computing* (pp. 11–23). Springer.

Baldridge, J., Morton, T., & Bierner, G. (2004). The opennlp maxent package in java. *URL: http://maxent. sourceforge. net*.

Banko, M., Cafarella, M. J., Soderland, S., Broadhead, M., & Etzioni, O. (2009). *Open information extraction for the web*. University of Washington.

Banko, M., Etzioni, O., & Center, T. (2008). The tradeoffs between open and traditional relation extraction. *Proceedings of ACL-08: HLT*, 28–36.

Bansal, M., Cardie, C., & Lee, L. (2008). The power of negative thinking: Exploiting label disagreement in the min-cut classification framework. *Proceedings of COLING: companion volume: Posters*, 13–16.

Bhattacharya, I., & Getoor, L. (2007). Collective entity resolution in relational data. *ACM Transactions on Knowledge Discovery from Data (TKDD)*, *1*(1), 5.

Biadsy, F., Hirschberg, J., Filatova, E., & InforSense, L. (2008). An unsupervised approach to biography production using wikipedia. In (pp. 807–815).

Bizer, C., Heath, T., & Berners-Lee, T. (2009). Linked data-the story so far. *International Journal on Semantic Web and Information Systems (IJSWIS)*, *5*(3), 1–22.

Blei, D. M., Ng, A. Y., & Jordan, M. I. (2003). Latent dirichlet allocation. *the Journal of machine Learning research*, *3*, 993–1022.

Blohm, S. (2010). *Large-scale pattern-based information extraction from the world wide web*. KIT Scientific Publishing.

Bollegala, D. T., Matsuo, Y., & Ishizuka, M. (2009). Measuring the similarity between implicit semantic relations from the web. In *Proceedings of the 18th international conference on world wide web* (pp. 651–660).

Brin, S. (1999). Extracting patterns and relations from the world wide web. In *The world wide web and databases* (pp. 172–183). Springer.

Broder, A. (2002). A taxonomy of web search. In *Acm sigir forum* (Vol. 36, pp. 3–10).

Broder, A., & Mitzenmacher, M. (2004). Network applications of bloom filters: A survey. *Internet Mathematics*, *1*(4), 485–509.

Brody, S., & Elhadad, N. (2010). An unsupervised aspect-sentiment model for online reviews. In *Human language technologies: The 2010 annual conference of the north american chapter of the association for computational linguistics* (pp. 804–812).

Bruce, R. F., & Wiebe, J. M. (1999). Recognizing subjectivity: a case study in manual tagging. *Natural Language Engineering*, *5*(2), 187–205.

Bunescu, R., & Mooney, R. (2006). Subsequence kernels for relation extraction. *Advances in neural information processing systems*, *18*, 171.

Bunescu, R., & Mooney, R. (2007). Learning to extract relations from the web using minimal supervision. In *Annual meeting-association for computational linguistics* (Vol. 45, p. 576).

Carlson, A., Betteridge, J., Kisiel, B., Settles, B., Hruschka Jr, E. R., & Mitchell, T. M. (2010). Toward an architecture for never-ending language learning. In *Proceedings of the twenty-fourth conference on artificial intelligence (aaai 2010)* (Vol. 2, pp. 3–3).

Chambers, N., & Jurafsky, D. (2008). Jointly combining implicit constraints improves temporal ordering. In *Proceedings of the conference on empirical methods in natural language processing* (pp. 698–706).

Chang, J., Boyd-Graber, J., & Blei, D. M. (2009). Connections between the lines: augmenting social networks with text. In *Proceedings of the 15th acm sigkdd international conference on knowledge discovery and data mining* (pp. 169–178).

Chaovalit, P., & Zhou, L. (2005). Movie review mining: A comparison between supervised and unsupervised classification approaches. In *System sciences, 2005. hicss'05. proceedings of the 38th annual hawaii international conference on* (pp. 112c–112c).

Ciaramita, M., Gangemi, A., Ratsch, E., Šarić, J., & Rojas, I. (2008). Unsupervised learning of semantic relations for molecular biology ontologies. In *Proceeding of the 2008 conference on ontology learning and population: Bridging the gap between text and knowledge* (pp. 91–104).

Clark, P., Harrison, P., & Thompson, J. (2003). A knowledge-driven approach to text meaning processing. In *Proceedings of the hlt-naacl 2003 workshop on text meaning-volume 9* (pp. 1–6).

Craven, M., DiPasquo, D., Freitag, D., McCallum, A., Mitchell, T., Nigam, K., & Slattery, S. (2000). Learning to construct knowledge bases from the world wide web. *Artificial Intelligence*, *118*(1), 69–113.

Cunningham, H. (2002). Gate: an architecture for development of robust hlt applications hamish cunningham, diana maynard, kalina bontcheva, valentin tablan department of computer science university of sheffield.

Dalvi, N., Kumar, R., Pang, B., Ramakrishnan, R., Tomkins, A., Bohannon, P., ... Merugu, S. (2009). A web of concepts. In *Proceedings of the twenty-eighth acm sigmod-sigact-sigart symposium on principles of database systems* (pp. 1–12).

d'Aquin, M., Sabou, M., Dzbor, M., Baldassarre, C., Gridinoc, L., Angeletou, S., & Motta, E. (2007). Watson: a gateway for the semantic web.

De Marneffe, M.-C., MacCartney, B., & Manning, C. D. (2006). Generating typed dependency parses from phrase structure parses. In *Proceedings of lrec* (Vol. 6, pp. 449–454).

De Marneffe, M.-C., & Manning, C. D. (2008). Stanford typed dependencies manual. *URL http://nlp. stanford. edu/software/dependencies_manual. pdf*.

Dhar, V., Chang, E., & Stern, L. N. (2007). Does chatter matter? the impact of user-generated content on music.

Dimitrova, M., Finn, A., Kushmerick, N., & Smyth, B. (2002). Web genre visualization. In *In proc. conference on human factors in.*

Ding, L., Finin, T., Joshi, A., Pan, R., Cost, R. S., Peng, Y., ... Sachs, J. (2004). Swoogle: a search and metadata engine for the semantic web. In *Proceedings of the thirteenth acm international conference on information and knowledge management* (pp. 652–659).

D Pierce, C., Booth, D., Ogbuji, C., Deaton, C., Blackstone, E., & Lenat, D.

(2012). Semanticdb: A semantic web infrastructure for clinical research and quality reporting. *Current Bioinformatics*, *7*(3), 267–277.

Dumais, S. T., Furnas, G. W., Landauer, T. K., Deerwester, S., & Harshman, R. (1988). Using latent semantic analysis to improve access to textual information. In *Proceedings of the sigchi conference on human factors in computing systems* (pp. 281–285).

Durbin, S. D., Richter, J. N., & Warner, D. (2003). A system for affective rating of texts. *KDD Wksp. on Operational Text Classification Systems (OTC-3)*.

Eguchi, K., & Lavrenko, V. (2006). Sentiment retrieval using generative models. In *Proceedings of the 2006 conference on empirical methods in natural language processing* (pp. 345–354).

Esuli, A., & Sebastiani, F. (2005). Determining the semantic orientation of terms through gloss classification. In *Proceedings of the 14th acm international conference on information and knowledge management* (pp. 617–624).

Etzioni, O., Cafarella, M., Downey, D., Kok, S., Popescu, A.-M., Shaked, T., … Yates, A. (2004). Web-scale information extraction in knowitall:(preliminary results). In *Proceedings of the 13th international conference on world wide web* (pp. 100–110).

Fader, A., Soderland, S., & Etzioni, O. (2011). Identifying relations for open information extraction. In *Proceedings of the conference on empirical methods in natural language processing* (pp. 1535–1545).

Ferrucci, D., & Lally, A. (2004). Uima: an architectural approach to unstructured information processing in the corporate research environment. *Natural Language Engineering*, *10*(3-4), 327–348.

Gamon, M., Aue, A., Corston-Oliver, S., & Ringger, E. (2005). Pulse: Mining customer opinions from free text. In *Advances in intelligent data analysis vi* (pp. 121–132). Springer.

Ganu, G., Elhadad, N., & Marian, A. (2009). Beyond the stars: Improving rating predictions using review text content. In *12th international workshop on the web and databases.*

Ghose, A., & Ipeirotis, P. G. (2011). Estimating the helpfulness and economic impact of product reviews: Mining text and reviewer characteristics. *Knowledge and Data Engineering, IEEE Transactions on*, *23*(10), 1498–1512.

Glance, N., Hurst, M., & Tomokiyo, T. (2004). Blogpulse: Automated trend discovery for weblogs. In *Www 2004 workshop on the weblogging ecosystem: Aggregation, analysis and dynamics* (Vol. 2004).

Grefenstette, G., Qu, Y., Shanahan, J. G., & Evans, D. A. (2004). Coupling niche browsers and affect analysis for an opinion mining application. In *Proceedings of riao* (Vol. 4, pp. 186–194).

Guo, J., Xu, G., Cheng, X., & Li, H. (2009). Named entity recognition in query. In *Proceedings of the 32nd international acm sigir conference on research and development in information retrieval* (pp. 267–274).

Hachey, B. (2009). Multi-document summarisation using generic relation extraction. In *Proceedings of the 2009 conference on empirical methods in natural language processing: Volume 1-volume 1* (pp. 420–429).

Harth, A., Hogan, A., Umbrich, J., & Decker, S. (2008). Building a semantic web search engine: challenges and solutions.

Heflin, J., Hendler, J., & Luke, S. (1999). Shoe: A knowledge representation language for internet applications.

Hillard, D., Ostendorf, M., & Shriberg, E. (2003). Detection of agreement vs. disagreement in meetings: Training with unlabeled data. In *Proceedings of the 2003 conference of the north american chapter of the association for computational linguistics on human language technology: companion volume of the proceedings of hlt-naacl 2003–short papers-volume 2* (pp. 34–36).

Hofmann, T. (1999). Probabilistic latent semantic indexing. In *Proceedings of the 22nd annual international acm sigir conference on research and development in information retrieval* (pp. 50–57).

Hu, M., & Liu, B. (2004). Mining opinion features in customer reviews. In *Proceedings of the national conference on artificial intelligence* (pp. 755–760).

In, M. (2010). Wikipedia: the free encyclopedia. *St. Petersburg (Florida): Wikipedia Foundation*, *27*(2006), 7–4.

Inkpen, D. Z., Feiguina, O., & Hirst, G. (2006). Generating more-positive and more-negative text. In *Computing attitude and affect in text: Theory and applications* (pp. 187–198). Springer.

Jansen, B. J., Booth, D. L., & Spink, A. (2008). Determining the informational, navigational, and transactional intent of web queries. *Information Processing & Management*, *44*(3), 1251–1266.

Ji, H., Favre, B., Lin, W.-P., Gillick, D., Hakkani-Tur, D., & Grishman, R. (2013). Open-domain multi-document summarization via information extraction: Challenges and prospects. In *Multi-source, multilingual information extraction and summarization* (pp. 177–201). Springer.

Kambhatla, N. (2004a). Combining lexical, syntactic, and semantic features with maximum entropy models for extracting relations. In *Proceedings of the acl 2004 on interactive poster and demonstration sessions* (p. 22).

Kambhatla, N. (2004b). Combining lexical, syntactic, and semantic features with maximum entropy models for extracting relations. In *Proceedings of the acl 2004 on interactive poster and demonstration sessions* (p. 22).

Kambhatla, N. (2004c). Combining lexical, syntactic, and semantic features with maximum entropy models for extracting relations. In *Proceedings of the acl 2004 on interactive poster and demonstration sessions* (p. 22).

Kessler, B., Numberg, G., & Schütze, H. (1997). Automatic detection of text genre.

In *Proceedings of the 35th annual meeting of the association for computational linguistics and eighth conference of the european chapter of the association for computational linguistics* (pp. 32–38).

Kobayashi, N., Iida, R., Inui, K., & Matsumoto, Y. (2006). Opinion mining as extraction of attribute-value relations. In *New frontiers in artificial intelligence* (pp. 470–481). Springer.

Koppel, M., & Schler, J. (2005). Using neutral examples for learning polarity. In *Proceedings of international joint conferences on artificial intelligence (ijcai)* (Vol. 2005, pp. 1616–1616).

Ku, L.-W., Liang, Y.-T., & Chen, H.-H. (2006). Opinion extraction, summarization and tracking in news and blog corpora. In *Proceedings of aaai-2006 spring symposium on computational approaches to analyzing weblogs* (Vol. 2001).

Kumar, R., & Tomkins, A. (2009). A characterization of online search behavior. *IEEE Data Engineering Bulletin*, *32*(2), 3–11.

Leskovec, J. (2008). *Dynamics of large networks*. ProQuest.

Lin, D. (2003). Dependency-based evaluation of minipar. In *Treebanks* (pp. 317–329). Springer.

Lin, D., & Pantel, P. (2001). Dirt@ sbt@ discovery of inference rules from text. In *Proceedings of the seventh acm sigkdd international conference on knowledge discovery and data mining* (pp. 323–328).

Liu, H., Lieberman, H., & Selker, T. (2003). A model of textual affect sensing using real-world knowledge. In *Proceedings of the 8th international conference on intelligent user interfaces* (pp. 125–132).

Lu, Y., Zhai, C., & Sundaresan, N. (2009). Rated aspect summarization of short comments. In *Proceedings of the 18th international conference on world wide web* (pp. 131–140).

Mani, I., Verhagen, M., Wellner, B., Lee, C. M., & Pustejovsky, J. (2006). Machine learning of temporal relations. In *Proceedings of the 21st international conference on computational linguistics and the 44th annual meeting of the association for computational linguistics* (pp. 753–760).

McDonald, R., Pereira, F., Kulick, S., Winters, S., Jin, Y., & White, P. (2005). Simple algorithms for complex relation extraction with applications to biomedical ie. In *Proceedings of the 43rd annual meeting on association for computational linguistics* (pp. 491–498).

Miller, G. A. (1995). Wordnet: a lexical database for english. *Communications of the ACM*, *38*(11), 39–41.

Min, B., Shi, S., Grishman, R., & Lin, C.-Y. (2012). Towards large-scale unsupervised relation extraction from the web. *International Journal on Semantic Web and Information Systems (IJSWIS)*, *8*(3), 1–23.

Moens, M.-F. (2007). Information extraction: The power of words and pictures. *Journal of Computing and Information Technology*, *15*(4), 295–303.

Mooney, R. (1999). Relational learning of pattern-match rules for information extraction. In *Proceedings of the sixteenth national conference on artificial intelligence* (pp. 328–334).

Mullen, T., & Malouf, R. (2006). A preliminary investigation into sentiment analysis of informal political discourse. In *Aaai symposium on computational approaches to analysing weblogs (aaai-caaw)* (pp. 159–162).

Naing, M. M., Lim, E. P., & Goh, D. H. L. (2010). Ontology-based web annotation framework for hyperlink structures.

Newcombe, H., Kennedy, J., Axford, S., & James, A. (1959). *Automatic linkage of vital records.*

Pang, B., & Lee, L. (2004). A sentimental education: Sentiment analysis using subjectivity summarization based on minimum cuts. In *Proceedings of the 42nd annual meeting on association for computational linguistics* (p. 271).

Pang, B., Lee, L., & Vaithyanathan, S. (2002). Thumbs up?: sentiment classification using machine learning techniques. In *Proceedings of the acl-02 conference on empirical methods in natural language processing-volume 10* (pp. 79–86).

Pantel, P., & Pennacchiotti, M. (2006). Espresso: Leveraging generic patterns for automatically harvesting semantic relations. In *Proceedings of the 21st international conference on computational linguistics and the 44th annual meeting of the association for computational linguistics* (pp. 113–120).

Pasula, H., Marthi, B., Milch, B., Russell, S., & Shpitser, I. (2002). Identity uncertainty and citation matching..

Pustejovsky, J., Hanks, P., Sauri, R., See, A., Gaizauskas, R., Setzer, A., ... others (2003). The timebank corpus. In *Corpus linguistics* (Vol. 2003, p. 40).

Qiu, G., Liu, B., Bu, J., & Chen, C. (2011). Opinion word expansion and target extraction through double propagation. *Computational linguistics*, *37*(1), 9–27.

Radev, D. R., & McKeown, K. R. (1998). Generating natural language summaries from multiple on-line sources. *Computational Linguistics*, *24*(3), 470–500.

Ramakrishnan, R., & Tomkins, A. (2007). Toward a peopleweb. *Computer*, *40*(8), 63–72.

Rauber, A., & Müller-Kögler, A. (2001). Integrating automatic genre analysis into digital libraries. In *Proceedings of the 1st acm/ieee-cs joint conference on digital libraries* (pp. 1–10).

Ravichandran, D., & Hovy, E. (2002). Learning surface text patterns for a question answering system. In *Proceedings of the 40th annual meeting on association for computational linguistics* (pp. 41–47).

Richardson, M., & Domingos, P. (2006). Markov logic networks. *Machine learning*, *62*(1-2), 107–136.

Riloff, E., Wiebe, J., & Wilson, T. (2003). Learning subjective nouns using extraction pattern bootstrapping. In *Proceedings of the seventh conference on natural language learning at hlt-naacl 2003-volume 4* (pp. 25–32).

Schoenmackers, S., Etzioni, O., & Weld, D. S. (2008). Scaling textual inference to the web. In *Proceedings of the conference on empirical methods in natural language processing* (pp. 79–88).

Schubert, L., & Tong, M. (2003). Extracting and evaluating general world knowledge from the brown corpus. In *Proceedings of the hlt-naacl 2003 workshop on text meaning-volume 9* (pp. 7–13).

Shen, W., Li, X., & Doan, A. (2005). Constraint-based entity matching. In *Proceedings of the national conference on artificial intelligence* (Vol. 20, p. 862).

Shinyama, Y., & Sekine, S. (2006). Preemptive information extraction using unrestricted relation discovery. In *Proceedings of the main conference on human language technology conference of the north american chapter of the association of computational linguistics* (pp. 304–311).

Singhal, A. (2012, May). Introducing the knowledge graph: things, not strings. In *Official google blog-http://googleblog.blogspot.com/2012/05/introducing-knowledge-graph-things-not.html.*

Singla, P., & Domingos, P. (2006). Entity resolution with markov logic. In *Data mining, 2006. icdm'06. sixth international conference on* (pp. 572–582).

Spertus, E. (1997). Smokey: Automatic recognition of hostile messages. In *Proceedings of the national conference on artificial intelligence* (pp. 1058–1065).

Srihari, R., & Li, W. (1999). *Information extraction supported question answering* (Tech. Rep.). DTIC Document.

Subasic, P., & Huettner, A. (2001). Affect analysis of text using fuzzy semantic typing. *Fuzzy Systems, IEEE Transactions on*, *9*(4), 483–496.

Suchanek, F. M., Kasneci, G., & Weikum, G. (2007). Yago: a core of semantic knowledge. In *Proceedings of the 16th international conference on world wide web* (pp. 697–706).

Syed, Z., & Finin, T. (2010). Unsupervised techniques for discovering ontology elements from wikipedia article links. In *Proceedings of the naacl hlt 2010 first international workshop on formalisms and methodology for learning by reading* (pp. 78–86).

Talukdar, P. P., Wijaya, D., & Mitchell, T. (2012). Acquiring temporal constraints between relations. In *Proceedings of the 21st acm international conference on information and knowledge management* (pp. 992–1001).

Titov, I., & McDonald, R. (2008). Modeling online reviews with multi-grain topic models. In *Proceedings of the 17th international conference on world wide web* (pp. 111–120).

Tunstall-Pedoe, W. (2010). True knowledge: Open-domain question answering using structured knowledge and inference. *AI Magazine*, *31*(3), 80–92.

Turney, P. D. (2002). Thumbs up or thumbs down?: semantic orientation applied to unsupervised classification of reviews. In *Proceedings of the 40th annual meeting on association for computational linguistics* (pp. 417–424).

Turney, P. D. (2006). Similarity of semantic relations. *Computational Linguistics*, *32*(3), 379–416.

Vanderwende, L., Banko, M., & Menezes, A. (2004). Event-centric summary generation. *Working notes of DUC*.

Vegnaduzzo, S. (2004). Acquisition of subjective adjectives with limited resources. In *Proceedings of the aaai spring symposium on exploring attitude and affect in text: Theories and applications, stanford, us.*

Wang, Y., Zhu, M., Qu, L., Spaniol, M., & Weikum, G. (2010). Timely yago:

harvesting, querying, and visualizing temporal knowledge from wikipedia. In *Proceedings of the 13th international conference on extending database technology* (pp. 697–700).

Weld, D. S., Hoffmann, R., & Wu, F. (2009). Using wikipedia to bootstrap open information extraction. *ACM SIGMOD Record*, *37*(4), 62–68.

Wiebe, J., Bruce, R., Bell, M., Martin, M., & Wilson, T. (2001). A corpus study of evaluative and speculative language. In *Proceedings of the second sigdial workshop on discourse and dialogue-volume 16* (pp. 1–10).

Wiebe, J., & Riloff, E. (2005). Creating subjective and objective sentence classifiers from unannotated texts. In *Computational linguistics and intelligent text processing* (pp. 486–497). Springer.

Wiebe, J. M. (2000). Learning subjective adjectives from corpora. In *Proceedings of the national conference on artificial intelligence* (pp. 735–741).

Wiebe, J. M., Bruce, R. F., & O'Hara, T. P. (1999). Development and use of a gold-standard data set for subjectivity classifications. In *Proceedings of the 37th annual meeting of the association for computational linguistics on computational linguistics* (pp. 246–253).

Wilson, T., Wiebe, J., & Hoffmann, P. (2009). Recognizing contextual polarity: An exploration of features for phrase-level sentiment analysis. *Computational linguistics*, *35*(3), 399–433.

Wu, F., & Weld, D. S. (2010). Open information extraction using wikipedia. In *Proceedings of the 48th annual meeting of the association for computational linguistics* (pp. 118–127).

Yan, Y., Okazaki, N., Matsuo, Y., Yang, Z., & Ishizuka, M. (2009). Unsupervised relation extraction by mining wikipedia texts using information from the web. In *Proceedings of the joint conference of the 47th annual meeting of the acl and the 4th international joint conference on natural language processing of* *the afnlp: Volume 2-volume 2* (pp. 1021–1029).

Yao, L., Haghighi, A., Riedel, S., & McCallum, A. (2011). Structured relation discovery using generative models. In *Proceedings of the conference on empirical methods in natural language processing* (pp. 1456–1466).

Yates, A., Cafarella, M., Banko, M., Etzioni, O., Broadhead, M., & Soderland, S. (2007). Textrunner: open information extraction on the web. In *Proceedings of human language technologies: The annual conference of the north american chapter of the association for computational linguistics: Demonstrations* (pp. 25–26).

Yoshikawa, K., Riedel, S., Asahara, M., & Matsumoto, Y. (2009). Jointly identifying temporal relations with markov logic. In *Proceedings of the joint conference of the 47th annual meeting of the acl and the 4th international joint conference on natural language processing of the afnlp: Volume 1-volume 1* (pp. 405–413).

Yu, H., & Hatzivassiloglou, V. (2003). Towards answering opinion questions: Separating facts from opinions and identifying the polarity of opinion sentences. In *Proceedings of the 2003 conference on empirical methods in natural language processing* (pp. 129–136).

Zelenko, D., Aone, C., & Richardella, A. (2003). Kernel methods for relation extraction. *The Journal of Machine Learning Research*, *3*, 1083–1106.

Zhao, H., & Ram, S. (2008). Entity matching across heterogeneous data sources: An approach based on constrained cascade generalization. *Data & Knowledge Engineering*, *66*(3), 368–381.

Zhao, S., & Grishman, R. (2005). Extracting relations with integrated information using kernel methods. In *Proceedings of the 43rd annual meeting on association for computational linguistics* (pp. 419–426).

Zhu, J., Nie, Z., Liu, X., Zhang, B., & Wen, J.-R. (2009). Statsnowball: a statistical approach to extracting entity relationships. In *Proceedings of the 18th international conference on world wide web* (pp. 101–110).

Appendix A

List of medical informatics websites that were crawled for research work in this dissertation

1. http://www.webmd.com
2. http://www.healthcentral.com
3. http://www.localhealth.com
4. http://www.cdc.gov/az
5. http://www.medicineonline.com
6. http://www.mayoclinic.com
7. http://health.yahoo.net/directory/health-channels
8. http://www.cnn.com/HEALTH
9. http://health.discovery.com
10. http://health.nih.gov
11. http://emedicine.medscape.com/
12. http://www.sharecare.com/

Appendix B

SPARQL Queries executed in DBPedia for Wikipedia Evaluation

1. PREFIX foaf: <http://xmlns.com/foaf/0.1/ >

   PREFIX dbpedia2: <http://dbpedia.org/property/ >

   PREFIX dbpedia: <23http://dbpedia.org/ >

   SELECT ?name ?person WHERE

   ?person dbpedia2:birthPlace <http://dbpedia.org/resource/Atlanta >.

   ?person foaf:name ?name .

   LIMIT 10000

Appendix C

Medical Sentences used for CREATE:Fact evaluation

1. RSV in older children and adults causes a cold.
2. It causes severe depression day time anxiety attacks, heart flutters.
3. It causes severe itching and a whitish discharge.
4. It causes severe muscle spasms, starting around the neck, chest, and back.
5. It causes severe pain and has a crippling effect on my joints and feet and legs.
6. Instead of testing for the bacteria that causes syphilis, the VDRL test checks for antibodies to this bacteria.
7. The test can detect antibodies to the germ that causes syphilis.
8. The organism that causes bejel belongs to the same family as the bacterium that causes syphilis, pinta and yaws and is known as treponema.
9. Syphilis tests detect antibodies to the bacterium that causes syphilis (Treponema pallidum) in blood, body fluid or tissue.
10. The bacteria that causes yaws is similar to the bacteria that causes syphilis, so someone who has yaws may test positive for syphilis.
11. The preferred treatment at all stages is penicillin, an antibiotic medication that can kill the organism that causes syphilis.
12. A progressive disorder of the central and autonomic nervous systems, it is characterized by orthostatic hypotension, which causes dizziness or fainting.
13. A drop in blood supply everywhere - including the brain - which causes dizziness, light headeness or a sudden loss of consciousness.
14. I also regularly give her pain killers [paracetemol and Scopex], to stop underlying infections and pain - like irritable bowl syndrome - which causes stress and is a major cause of her psychotic episodes.
15. Abusive environments make the speech impairments even worse because the necessary patience and support for a child with a speech impairment is absent, which causes stress, frustration, and fear for a child.
16. That is not enough to last until my next dr appointment so I go through withdrawal on top of the pain which causes stress which effects my diabetes.
17. In 1984, scientists proved that HIV causes AIDS.

18. Although HIV causes AIDS, a person can be infected with HIV for many years before AIDS develops.
19. HIV causes AIDS (acquired immunodeficiency syndrome).
20. The best way to prevent urticaria is to avoid whatever causes it.
21. In fact, normal grief can become complicated by a depression, which in turn can interfere with the resolution of the grief; so, in other words, depression is a syndrome involving many components, and whatever causes it (we don't really know), it should get treated.
22. One key way to manage urticaria on a daily basis is to avoid whatever causes it.
23. Besides medications to manage symptoms, the best treatment for urticaria is to avoid whatever causes it.
24. Genital warts, also known as human papillomavirus (HPV), are a viral infection that causes small, hard painless bumps in the vaginal area or on the penis.
25. Oral herpes (herpes simplex labialis) is a very common disease that causes small, painful cold sores and fever blisters of the mouth, lips or gums.
26. Kissing bugs can carry a parasite that causes Chagas disease, but this is not common in the United States.
27. Currently, most of the U.S. blood supply is screened for Trypanosoma cruzi (the parasite that causes Chagas disease).
28. It is important to note that not all triatomine bugs are infected with the parasite that causes Chagas disease.
29. While it is not fully known what causes GBS, it is known that about two-thirds of people who get GBS do so several days or weeks after they have been sick with diarrhea or a lung or sinus illness.
30. Scientists do not fully understand what causes GBS, but it is believed that stimulation of the bodys immune system may play a role in its development.
31. Researchers from Massachusetts General Hospital and the University of Michigan have discovered a new clue in the mystery of what causes fibromyalgia pain.
32. While there are many theories on what causes fibromyalgia, the truth remains that scientists do not know.
33. Doctors don't know exactly what causes fibromyalgia.

34. While there are many theories on what causes fibromyalgia, there is no known cause or cure.

35. While doctors don't know exactly what causes fibromyalgia, they do know that fibromyalgia sufferers have an imbalance of certain brain chemicals.

36. However, at this point there is no widely held theory for what causes fibromyalgia.

37. Unfortunately, without knowing specifically what causes fibromyalgia, there's really no way of knowing.

38. Its not known what causes fibromyalgia.

39. Hi smtip 10 We don't really know exactly what causes fibromyalgia but since it tends to recur in families, many researchers think some of us have a genetic predisposition to FM that may be triggered by some type of trauma like an injury or illness.

40. There has been intense debate about whether thimerosal causes autism, a link repeatedly discounted in scientific studies.

41. The new study from researchers at the University of Missouri-Columbia adds to a growing body of evidence refuting claims that thimerosal causes autism, study author Judith Miles, M.D., Ph.D., told Ivanhoe.

42. Based on these studies, they concluded that there was no evidence that thimerosal causes autism.

43. I think the problem originates in the brain because of the way it affects the same part of the visual field in both eyes, but it is far less severe than a typical migraine since it causes no pain.

44. Astigmatism may be so slight that it causes no problems.

45. In some cases, it causes no symptoms.

46. This causes more normal shaped cells that do not impede blood flow.

47. This causes more calcium absorption, which increases the calcium level in your blood.

48. Once the ameba enters the nose, it travels to the brain where it causes PAM, which is usually fatal.

49. I am getting a lot more nerve pain since reducing the patch but I am determined to get off it now, mostly due to the severe constipation it causes which i absolutely hate.

50. We do not know what causes dyslexia, but we do know that it affects children who are physically and emotionally healthy, academically capable, and who come from good home environments.
51. Experts do not know precisely what causes dyslexia, but several recent studies now indicate that genetics plays a major role.
52. Whatever causes it, the reaction continues without control and damages the intestinal wall, leading to diarrhea and abdominal pain.
53. Whatever causes it, fatigue usually goes away during the second trimester.
54. It is not known what causes these symptoms.
55. Experts still aren't sure what causes these headaches.
56. Is there treatment and what causes these spasms?
57. I always end up at the er but nobody can tell me what causes these bouts.
58. The answer could shed light on what causes these conditions.
59. find out what causes these things?
60. Researchers are unsure what causes these benign tumors, though they speculate it has something to do with excess tissue left over from early pituitary development.
61. Researchers don't know what causes these genetic mutations.
62. Good luck - hope you can figure out what causes these lumps, and get them treated.
63. It's unclear what causes these irregularities, but ADHD runs in families, so many experts believe genetics play a role.
64. No one knows what causes these small, painful blisters inside your mouth.
65. I have no idea what causes these changes, I try to stay the same everytime for each reading.
66. Determining and then avoiding what causes these bodily reactions can prove helpful.
67. Doctors do not know what causes these autoimmune flares but do know that they do not have to be triggered by gluten.
68. How it interrupts sleep: Doctors don't know exactly what causes these sleep movement disorders, but they do know they're directly related to a lack of deep, restful, REM sleep.

69. To find out what causes these abnormalities, Mantzoros' group looked at leptin levels in female athletes with hypothalamic amenorrhea.

70. Exactly what causes acne?

71. No one is sure what causes acne, but hormones appear to play a supporting role.

72. Primary hemochromatosis is usually caused by a specific genetic problem that causes too much iron to be absorbed.

73. It results in the bone marrow making an enzyme, called tyrosine kinase, that causes too many stem cells to develop into white blood cells (granulocytes or blasts).

74. Tetanus (also known as lockjaw) is a very serious illness that causes convulsions (seizures) and severe muscle spasms that can be strong enough to cause bone fractures of the spine.

75. Tetanus is a serious illness that causes convulsions (seizures) and severe muscle spasms that can be strong enough to cause bone fractures of the spine.

76. Cancer screening trials also are meant to show whether early detection (finding cancer before it causes symptoms) decreases a person's chance of dying from the disease.

77. Because diabetic retinopathy can begin and get a foothold before it causes symptoms, all patients with diabetes should have an eye examination with pupils dilated at least once a year.

78. Most doctors consider blood pressure to be too low when it causes symptoms or drops suddenly.

79. Instead, blood pressure is considered too low if it causes symptoms.

80. There are two ways to screen for ovarian cancer before it causes symptoms or shows up during a routine gynecologic exam.

81. That is because in a majority of patients, when acid refluxes up into the esophagus from the stomach, it causes symptoms, but when the acid returns back to the stomach it does not cause any damage (or esophagitis).

82. However, when it causes symptoms, it can be described as extremely painful.

83. You may not know you have antiphospholipid syndrome (APS) until it causes symptoms.

84. Conversely, a disease may become the focus of preventive medicine despite its low prevalence because it causes significant illness, disability, and death (e.g., a disease with a high case-fatality rate such as infection with the Ebola virus).

85. This will lead to better control of RA before it causes significant disability.
86. However, treatment may be necessary if gynecomastia doesn't improve on its own, or if it causes significant pain, tenderness or embarrassment.
87. Others have celiac disease and it causes intestinal pain and bloating.
88. A sciatic hernia can be life-threatening if it causes intestinal blockage or becomes incarcerated and strangulated.
89. When it happens in the arteries to the intestine, it causes intestinal ischemia.
90. An obturator hernia is a rare but life-threatening type of abdominal hernia when it causes intestinal blockage or becomes incarcerated and strangulated.
91. Chances are good that if a parent has a food or drug allergy that causes hives, the child will experience a similar reaction.
92. Anyway, I furthered my research to find out "why me"...and discovered that we special folks emit an excess amount of histamine (same thing that causes hives) during the healing process.
93. I have a food related allergy that causes hives when i eat bagels or croissants, but not when i eat bread.
94. Shingles is a disease caused by the same virus that causes chickenpox, the varicella zoster virus.
95. Seek prompt medical care if you, or someone you are with, have symptoms of shingles, particularly the following symptoms or conditions: Shingles is a painful disease caused by the same virus that causes chickenpox, the varicella zoster virus.
96. In postherpetic neuralgia, the nerve inflammation is caused by a recent infection by the varicella- zoster virus, the same virus that causes chickenpox.
97. It sounds like you have a recuring case of the shingles..shingles or herpes zoster is the same virus that causes chickenpox..it can lay dormant in your body for years..it normally effects one side of the body..symptoms usually last for about 3 weeks..it is rare that the symptoms reoccur but it is known to happen in some cases.
98. Viruses such as herpes simplex, varicella-zoster (a virus that causes chickenpox and shingles), and Epstein-Barr damage sensory nerves and cause intense episodes of shooting pain.
99. Shingles is an infection caused by the virus varicella-zoster, which is the same virus that causes chickenpox.
100. Ramsay Hunt syndrome is caused by the same virus that causes chickenpox.

101. However, many other types of viruses, such as the herpes simplex virus, the mumps and measles viruses (against which most children are protected due to mass immunization programs), the virus that causes chickenpox, the rabies virus, and a number of viruses that are acquired through the bites of infected mosquitoes.

102. Shingles is a painful localized skin rash often with blisters that is caused by the varicella zoster virus (VZV), the same virus that causes chickenpox.

103. The virus that causes shingles varicella-zoster virus is also the virus that causes chickenpox.

104. Shingles is caused by the varicella zoster virus, the same virus that causes chickenpox.

105. A variant of Bell's palsy, called Ramsay-Hunt syndrome, is caused by the herpes zoster virus, the virus that causes chickenpox and shingles.

106. Shingles is caused by the same virus that causes chickenpox, and, thus, is in the vaccine.

107. Shingles is caused by the varicella-zoster virus the same virus that causes chickenpox.

108. Shingles is caused by the same herpes virus that causes chickenpox.

109. Shingles develops from the virus that causes chickenpox (varicella-zoster virus).

110. Shingles is a disease that is brought about by the virus that causes chickenpox.

111. It is the same virus that causes chickenpox and shingles, a disease of the nervous system.

112. This infection is produced by varicella zoster, the virus that causes chickenpox.

113. Most people are surprised to find out that two types of herpes viruses – the one that causes cold sores and the one that causes chickenpox – can cause a condition called herpetic eye disease.

114. Varicella-zoster is the virus that causes chickenpox.

115. About half of all cases occur among men and women who are 60 years old or older... more Shingles is caused by the varicella zoster virus, the same virus that causes chickenpox.

116. Shingles is a disease caused by the same virus that causes chickenpox.

117. Herpes viruses include varicella zoster (the virus that causes chickenpox), Epstein-Barr virus (the virus that causes mononucleosis), herpes simplex 1 (the virus that causes cold sores), and herpes simplex 2 (the virus that causes genital herpes).

118. But the virus that causes chickenpox and shingles is not the same virus responsible for cold sores or genital herpes, a sexually transmitted infection.

119. The disease, which is caused by the varicella zoster virus (the same virus that causes chickenpox) results in a painful rash that leads to blisters and scabs within 7 to 10 days.

120. The varicella-zoster virus that causes chickenpox can, much later in life, produce painful nerve conditions such as shingles and postherpetic neuralgia.

121. The virus that causes shingles, the varicella-zoster virus, is the same virus that causes chickenpox.

122. Tests for exposure to diseases such as toxoplasmosis and varicella (the virus that causes chickenpox) may also be done if needed.

123. The virus that causes chickenpox, the varicella zoster virus (VSV), can become dormant in nerve cells after an episode of chickenpox and later reemerge as shingles.

124. VariZIG is a product recommended for people who do not have immunity to varicella (chickenpox), who are at high risk for severe chickenpox or complications, and who are ineligible to receive the chickenpox vaccine to protect them from getting chickenpox or to attenuate the severity of disease after being exposed to the virus that causes chickenpox.

125. Shingles is a painful skin rash caused by the varicella zoster virusthe same virus that causes chickenpox.

126. Shingles is an infection caused by the varicella-zoster virus (VZV), which is the virus that causes chickenpox.

127. Varicella is the virus that causes chickenpox.

128. Shingles, also known as zoster or herpes zoster, is a painful skin rash caused by the varicella zoster virus, the same virus that causes chickenpox.

129. This disease is caused by the varicella zoster virus (the same virus that causes chickenpox).

130. Shingles is an infection caused by the same virus that causes chickenpox, the varicella-zoster virus.

131. Shingles is caused by the same virus that causes chickenpox.

132. Shingles occurs when the virus that causes chickenpox starts up again in your body.

133. Infectious agents suspected of causing TM include varicella zoster (the virus that causes chickenpox and shingles), herpes simplex, cytomegalovirus, Epstein-Barr, influenza, echovirus, human immunodeficiency virus (HIV), hepatitis A, and rubella.

134. When the virus that causes chickenpox reactivates, it causes shingles.

135. Shingles (herpes zoster) is a painful skin rash, often with blisters that's caused by the varicella-zoster virus, the same virus that causes chickenpox.

136. Zoster is caused by the same virus that causes chickenpox.

137. If you had chicken pox it might be Shingles (herpes zoster) is a painful, blistering skin rash due to the varicella-zoster virus, the virus that causes chickenpox.

138. A few specific vaccinations, such as the MMR (measles-mumps-rubella), varicella (the virus that causes chickenpox), or hepatitis A vaccines increase the risk of birth defects.

139. Shingles, medically called herpes zoster, is a painful skin rash that's caused by the same virus that causes chickenpox.

140. Shingles is a reactivation of the varicella-zoster virus, a type of herpes virus that causes chickenpox.

141. When one of these genes is damaged in humans, it causes deafness, according to Cargill.

142. (Streptomycin) It doesn't mean the antibotic is bad, it means in some people it causes deafness.

143. Asthma is a chronic inflammatory disease of the respiratory system that causes breathing difficulty.

144. Asthma is a disease of the respiratory system that causes breathing difficulty.

145. This is a potentially life-threatening condition that causes breathing difficulties and possible cardiac arrest.

146. Asthma is a chronic (long-term) lung condition that causes breathing difficulties and wheezing when air passages narrow and become inflamed.

147. "Parkinsonism" refers to any condition that causes Parkinson's-type abnormal movements.

148. The two drugs use different mechanisms to counteract the decline in the production of dopamine in the brain that causes Parkinson's symptoms.

149. No known treatment can stop or reverse the breakdown of nerve cells that causes Parkinson's disease.

150. Researchers think that shouting and lashing out during sleep may be caused by the same brain malfunction that causes Parkinson's and other types of dementia.

151. The two drugs use different mechanisms to counteract the decline in the production of dopamine in the brain that causes Parkinson"s symptoms.

152. And when given a small dose of a chemical that causes Parkinson's disease, the mice on the low-folic-acid diet developed muscle coordination problems as seen in people with Parkinson's.

153. Plaque psoriasis is a long-term disease that causes red scaly patches, often with a silvery scale, to appear on the skin.

154. Eczema is an allergic-type condition that causes red, irritated, and itchy skin.

155. Doctors do not know what causes chronic fatigue syndrome (CFS).

156. It is not known what causes chronic fatigue syndrome.

157. "There is no single test that will determine what causes chronic cough.

158. Experts don't fully understand what causes chronic rejection.

159. Unfortunately, by not clearly understanding what causes chronic itching, treatment may not be effective.

160. Many scientific studies have linked chronic fatigue syndrome to viral infections, such as infections with the Epstein-Barr virus, but a definitive association has not been established.... Read more about chronic fatigue syndrome introduction It is not known what causes chronic fatigue syndrome.

161. You never have to accept behavior which causes you harm.

162. Insomnia often leads to sleep deprivation, which causes you not only to be tired but also to feel irritable and lethargic and have difficulty focusing on tasks.

163. Get the support that you need to cope at the job you have or to help you to generate ideas of how to find a job which causes you less stress.

164. Shingles occurs when the varicella zoster virus, which causes chickenpox, is reactivated.

165. This group of viruses includes the herpes simplex viruses, varicella-zoster virus (which causes chickenpox and shingles), and Epstein-Barr virus (which causes infectious mononucleosis, also known as mono).

166. Herpes zoster, or shingles, is a reactivation of the varicella virus, which causes chickenpox.

167. Because science still does not know what causes a pituitary tumor, there are no surefire ways to prevent them at this time.

168. Your doctor will put different substances on your skin to see what causes a local reaction.

169. Understand what causes a high-risk pregnancy, and what you can do to take care of yourself and your baby.

170. Coronary Artery Disease or CAD is what causes a heart attack.

171. Scientists are not sure what causes a pituitary tumor.

172. It isn't clear what causes a meningioma.

173. By understanding what causes a weight-loss plateau, you can decide how to respond and avoid backsliding on your healthy-eating and exercise habits.

174. Occupational studies of workers exposed to high levels of benzene have shown that benzene causes leukemia, a cancer of the bone marrow (where blood cells are made).

175. Previous research carried out by this investigative team and by other scientists has shown that benzene causes leukemia and lowers blood cell counts in people who are exposed to high levels of benzene at work.

176. Fibromyalgia is a condition that causes widespread muscle and soft tissue pain and tenderness, especially in the trunk, neck, and shoulders.

177. Savella is a proven effective medication for the management of fibromyalgia - a common, chronic condition that causes widespread pain and affects an estimated 6-12 million people in the United States alone.

178. Even though they afflict millions of persons around the world, several of the common musculoskeletal disorders fall into the category of moderately prevalent, including gout, a form of episodic arthritis; fibromyalgia, a disorder of diffuse muscular pain and a subtype of soft tissue rheumatism; and rheumatoid arthritis, an inflammatory systemic disorder that causes widespread joint pain.

179. FM is a hypersensitivity disorder that causes widespread pain, fatigue, poor sleep, problems thinking etc.

180. In addition, duloxetine is used to help relieve nerve pain (peripheral neuropathy) in people with diabetes or ongoing pain due to medical conditions such as arthritis, chronic back pain, or fibromyalgia (a condition that causes widespread pain).

181. Since most urethritis is caused by sexually transmitted infections, your doctor will examine you for signs of other infections including syphilis, human papilloma virus (HPV) that causes venereal warts and HIV.

182. Recurrent vaginal yeast infections, chronic pelvic inflammatory disease, frequently recurring severe genital herpes, or human papillomavirus (HPV) – the virus that causes venereal warts (condyloma) – can also indicate that HIV infection has progressed to AIDS.

183. Botox Cosmetic is a sterile, purified version of the same toxin that causes botulism, a severe form of foodborne illness.

184. The agent that causes botulism (Clostridium botulinum), for example, originates in soil, but the source of most botulism infections is improperly canned food containing C. botulinum spores.

185. Botox, which are injections of a certain toxin that causes botulism, is a popular choice in eliminating wrinkles, but it's also possible that it could eliminate urinary incontinence as well.

186. Botox (botulinum toxin) is a protein that causes botulism, a serious and sometimes fatal foodborne illness.

187. "Currently our jerky has the minimum amount of sodium nitrate necessary to prevent the growth of the bacteria that causes botulism.

188. Botox is a protein that's derived from the bacteria that causes botulism, a severe form of food poisoning.

189. Likewise, accidental food contamination by botulinum toxin (the agent that causes botulism), E. coli, and other harmful organisms during the storage or preparation of food is much more likely than intentional food poisoning.

190. A psychiatrist, especially one who specializes in child and adolescent issues, is a good person to talk to.

191. In some cases when you call to set up an appointment, you may be referred immediately to a specialist, such as a pediatric psychiatrist or other mental health provider who specializes in child development.

192. If a biopsy indicates that you have cancer, you should consult an oncologist, a doctor who specializes in cancer, as soon as possible.

193. Your doctor can also refer you to an outside counselor in the area who specializes in cancer support.

194. Most likely, if nasopharyngeal cancer is suspected, you will be sent to a doctor who specializes in the ears, nose, and throat, or a doctor who specializes in cancer.

195. Depending on your situation and the outcome of any tests, you may be referred to a doctor who specializes in skin diseases (dermatologist) or to a doctor who specializes in cancer treatment (oncologist).

196. If the biopsy is positive for cancer, be sure to seek a confirming opinion by a doctor who specializes in cancer treatment before any treatment is started.

197. Whether you undergo silicone gel reconstruction, free flap or autologous tissue transfer, a board certified plastic and reconstructive surgeon who specializes in cancer reconstruction of the breast will have the experience to give you the results you are looking for.

198. An oncologist is a physician who specializes in cancer care.

199. (A doctor who specializes in cancer treatment is known as an oncologist.)

200. It makes sense that a surgeon who specializes in cancer surgery would be more proficient at it, since (s)he deals only with cancer cases; and at least one study has shown that surgical oncologists get more successful results than general surgeons when it comes to dealing with cancer surgery.

201. Oncologist: This is a doctor who specializes in cancer treatment.

202. Please go to a doctor who understands chronic pain, who specializes in it, rather than just having it as a sideline.

203. If your primary care doctor can exclude everything but TMJ its time to find a doctor or dentist who specializes in it.

204. You can apply online for social security disability and even look for a lawyer who specializes in it (since it can be a bit complex).

205. You can bypass your doctor if you feel he doesn't understand adult ADHD and instead, find a medical professional who specializes in it.

206. I'm not saying she has chronic fatigue but a doctor who specializes in it might know which tests to run to rule out the causes of the fatigue.

207. Consult an Eye M.D. If you still don't get a satisfactory answer you can consult a neuroopthalmologist who is an Eye M.D. that specializes in neurological conditions that affect the eye.

208. through some folks at my MS meeting I found a clinic that specializes in neurological diseases, I had an appointement with him and he went over ALL my MRI's from Jan 09 till current, as of then. he then came into the office and refered back to my chart several times.

209. I too have brain fog and I have had terrible insomnia and was sent to a nuero doc who specializes in sleep disorders.

210. Patients who have chronic insomnia or other severe sleep disturbances (particularly sleep apnea) may want to consult a doctor who specializes in sleep disorders.

211. That is the first he sleept that way in years.Need info on where to go for help with this.Marge 561-762-5211 Have you considered going to a dentist who specializes in sleep disorders?

212. These therapies are most often administered by a psychologist who specializes in sleep disorders.

213. Talk to your doc, or find a new one if needed, and maybe see about getting a counselor or therapist who specializes in addiction.

214. "Poker is the new rage among adolescents, and kids as young as 9 are now playing," says JoAnn White, PhD, a therapist who specializes in addiction in Cherry Hill, N.J. "More than 8% of new gamblers may end up having some type of gambling addiction, but we don't know how to identify then in advance," White says.

215. This will assist you in finding a dietitian that specializes in your specific needs.

216. Please consult an a physician that specializes in your particular concerns so that you may receive the best care possible.

217. In most cases, the diagnosis will be made by a family doctor or a urologist, a doctor who specializes in male reproductive disorders and urinary tract problems.

218. Your doctor may refer you to a specialist such as a doctor who specializes in male genital problems (urologist), a doctor who specializes in the hormonal systems (endocrinologist), a doctor who diagnoses and treats mental health problems (psychiatrist), or another type of specialist.

219. Depending on your particular health concerns, you may go directly to a specialist such as a doctor who specializes in male genital problems (urologist) or a doctor who specializes in the hormonal systems (endocrinologist).

220. Working with a therapist who specializes in chronic illnesses can help your child learn coping strategies.

221. Seeking out a counselor who specializes in chronic illness might be the way to go.

222. The other thing I highly recommend is finding a psychologist who specializes in chronic pain.

223. Ms. Rodrguez is a nurse who specializes in chronic and critical care, and is a certified diabetes educator.

224. I now have an appointment Oct. 4. Another month to wait, but no doubt worth it if it means I'll be seeing someone who specializes in chronic headache and migraine.

225. Be open with your partner about your concerns, and consider seeing a therapist who specializes in chronic illness.

226. He was instructed to speak to the reindeer counselor who specializes in chronic illness to learn how to cope with Migraine - something that is mandatory at Santa's company.

227. Finally, go in search of support from other people dealing with chronic illness - either online or in a support group - or find a counselor who specializes in chronic illness.

228. who specializes in chronic back pain :(, seems your family pphysician can not prescribe long term pain meds anymore ?! wish you luck...if you are in a lot of pain and what kind of pain?

229. You might also discuss with your parents the possibility of you talking with a psychologist or therapist who specializes in chronic pain issues.

230. This is a form of plastic surgery, and is often (but not necessarily) done by a periodontist who specializes in gum treatments.

231. Treatment options for periodontitis begins when your dentist, periodontist (a dentist who specializes in gum disease), or dental hygienist thoroughly removes the plaque and tartar that has built up on your teeth and caused the infection.

232. He or she may be a family practice doctor or an internist – someone who specializes in internal medicine and the study of disease in adults.

233. A physician who specializes in internal medicine is referred to as an internist.

234. Demonstrate your inhaler technique to a licensed Respiratory Therapist or a nurse who specializes in COPD and ask if you are using good technique.

235. The association recommends that people with COPD work with a registered dietitian who specializes in COPD in order to develop a food plan, find information on food-related issues (such as what to opt for when eating in

restaurants), get recommendations about cooking resources, and identify any potential drug-food interactions that may result from the medications being taken.

236. You don't mention what kind of doctor gave you the inhaler, but I would strongly suggest that you locate a pulmonologist who specializes in COPD - he/she would have much more detailed experience and knowledge of the disease and treatment than a GP.

237. Husband found a Migraine clinic with a new neurologist that specializes in headaches/migraines only.

238. Husband found a Migraine clinic with a new neurologist that specializes in headaches/migraines only.

239. Lifestyle changes and medications may be a first step all by themselves - or they may be combined with a minimally invasive clinical procedure performed by a physician who specializes in the health of blood vessels.

240. A prostate biopsy is done by a urologist, a doctor who specializes in the urinary system and men's sex organs.

241. However, in some cases when you call to set up an appointment, you may be referred immediately to a doctor who specializes in the treatment of hormone-related conditions in children (pediatric endocrinologist).

242. If you suspect you have Chronic Migraine, please visit RewriteYourDay.com to find a doctor who specializes in the diagnosis and management of the condition.

243. Or, when you call to set up an appointment, you may be referred to a doctor who specializes in the diagnosis and treatment of heart conditions (cardiologist).

244. However, if your child likely has reactive attachment disorder or another mental health problem, you'll need to see a doctor who specializes in the diagnosis and treatment of mental illness (psychiatrist) for a complete evaluation.

245. The best we can do is being sure to gather the best information available and work with a provider who specializes in the art and science of managing menopause.

246. You might do well to find a therapist who can teach you self tractioning and mobilization techniques (maybe even a chiropractor who specializes in the upper cervical spine).

247. The tissue will be examined by a pathologist, a physician who specializes in the diagnosis of diseased tissues.

248. You might think about seeing a pulmonologist, who specializes in the care of asthmatic patients.

249. If your doctor suspects you may have a liver problem, such as Wilson's disease, you may be referred to a doctor who specializes in the liver (hepatologist).

250. Ophthalmologists are physicians who specializes in the medical and surgical care of the eyes and visual system and in the prevention of eye disease and injury.

251. A doctor who specializes in the condition should monitor a pregnancy complicated by HSV infection.

252. However, in some cases when you call to set up an appointment, you may be referred to a doctor who specializes in the diagnosis and treatment of bleeding disorders (hematologist).

253. If your doctor suspects you may have solitary rectal ulcer syndrome, you may be referred to a doctor who specializes in the digestive system (gastroenterologist).

254. (Mayo Clinic) Your urologista doctor who specializes in the urinary tractmay also choose to inject collagen directly into the supportive tissues of your urethra.

255. PathologistA doctor who specializes in the anatomic (structural) and chemical changes that occur with diseases.

256. A neonatologist is a pediatric doctor who specializes in the diagnosis and treatment of disorders in newborns.

257. You may be referred to a cardiologist a doctor who specializes in the study of the heart and its function for tests such as:

258. Find a doctor who specializes in the area you are having symptoms with before the syptoms kill you.

259. After an initial examination, it's likely that the doctor will refer you to a doctor who specializes in the diagnosis and treatment of heart conditions (cardiologist) or a heart rhythm specialist (electrophysiologist).

260. After an initial examination, your doctor may refer you to an ear, nose and throat (ENT) specialist or a doctor who specializes in the brain and nervous system (neurologist).

261. After your family doctor or your child's pediatrician evaluates your child, your child may be referred to a doctor who specializes in the diagnosis and treatment of conditions related to the adrenal glands (endocrinologist).

262. A friend recommended I see her cardiologist, and within hours he sent me to an electrophysiologist, a doctor who specializes in the electrical activity of the heart.

263. In some cases when you call to set up an appointment, you may be referred to a doctor who specializes in the diagnosis and treatment of heart conditions (cardiologist).

264. Because there are more than 100 different types of cancer, finding someone who specializes in the type of cancer you have may be a challenge.

265. The team of doctors overseeing treatment can include a neurosurgeon who specializes in the brain and nervous systems, an oncologist, a radiation oncologist who practice radiation therapy, and your primary health care provider.

266. Someone who specializes in the gastro tract, stomach and esophagus.

267. I don't know if you have the same kind of specialist doctors in Egypt, but her in the U.S., I would suggest that someone with your symptoms see a neurologist, who specializes in the central nervous system the brain and spinal cord.

268. Orthodontic treatment may be provided by your dentist or an orthodontist, a dentist who specializes in the diagnosis, prevention and treatment of dental and facial irregularities.

269. This is a doctor who specializes in the treatment of lung conditions such as COPD and asthma.

270. Most likely, if nasopharyngeal cancer is suspected, you will be sent to a doctor who specializes in the ears, nose, and throat, or a doctor who specializes in cancer.

271. A physician who specializes in the care of frail elders, or those with complicated medical needs, is known as a geriatrician.

272. After an initial evaluation, your doctor may refer you to a doctor who specializes in the diagnosis and treatment of conditions that affect the brain and nervous system (neurologist).

273. Skip tough love and chooe the best professional, almost aways , a psychiatrist, who specializes in the treatment of these disorders, especially AADHD , and respects the needs of children and adults who have suffered for years being told that they could, if they just wanted to.

274. These stories will also help other Migraineurs see that theyre not alone and encourage those who think they have Chronic Migraine to seek care from a doctor who specializes in the diagnosis and management of the condition.

275. A rheumatologist is a physician who specializes in the treatment of joint disorders and arthritis.

276. An ophthalmologist is a healthcare professional with a Doctor of Medicine (MD) degree who specializes in the treatment of the eyes and related structures.

277. The patient may wish to consider consulting a sleep clinic or a doctor who specializes in the treatment of sleep disorders as well as their family doctor.

278. In some cases, however, you may be referred to a doctor who specializes in the diagnosis and treatment of skin conditions (dermatologist).

279. A pathologist is a physician who specializes in the diagnosis and development of diseases.

280. A neurologist is a doctor who specializes in the diagnosis and treatment of disorders of the nervous system.

281. Your regular doctor may refer you to a rheumatologist, a doctor who specializes in the treatment of osteoarthritis and other inflammatory conditions of the joints, muscles, and bones.

282. If your doctor suspects you may have gallstones, you may be referred to a doctor who specializes in the digestive system (gastroenterologist) or to an abdominal surgeon.

283. Dentists may further specialize in areas such as: A dentist is a doctor who specializes in the oral cavity to diagnose and treat diseases of the teeth and gums.

284. A pathologist, a physician who specializes in the study of diseased tissue, examines the tumor samples to identify the cancer type and stage.

285. A person with a severe case of viral hepatitis may need to see a a doctor who specializes in the digestive system (a gastroenterologist) and may require hospital treatment.

286. The sample is examined under a microscope by a doctor who specializes in the effects of disease on body tissues (a pathologist) to detect abnormalities of the liver.

287. Dr. Peek-Asa is an epidemiologist who specializes in the evaluation of injury prevention programs.

288. An implantable loop recorder is inserted by an electrophysiologist, a physician who specializes in the hearts electrical system.

289. Your pathology report is compiled by a pathologist, a doctor who specializes in the microscopic examination of tissue to identify abnormalities.

290. Its important to identify a cardiologist who specializes in the area in which you need them, says Dr. Foody.

291. After asking a few professionals in a few different fields, I learned of a gyno who specializes in the hormone area, as well as other well-being approaches.

292. If I were in your position, I'd look for a therapist who specializes in the problems of teenagers - this is an emotional shock, and treating physical symptoms won't help, as you well know.

293. If you are considering these or other dietary changes, talk to both a doctor who specializes in the digestive system (gastroenterologist) and a registered dietitian.

294. Internist: A physician who specializes in the diagnosis and medical treatment of adults.

295. A radiologist is a doctor who specializes in the branch of medicine that uses imaging technologies such as X-ray, CT scans, and MRIs to diagnose and treat diseases.

296. Dr. Smith is a Mayo Clinic neuropsychologist who specializes in the treatment of Alzheimer's disease.

297. In some cases, you might need to see a specialist, such as an otolaryngologist, who specializes in the ear, nose, and throat.

298. Oncologistmedical, radiation, and surgical: A medical oncologist is a doctor who specializes in the treatment of cancer.

299. If your doctor suspects you have cholecystitis, you may be referred to a doctor who specializes in the digestive system (gastroenterologist), or your doctor may refer you directly to the hospital.

300. You may be referred to a cardiologist a doctor who specializes in the study of the heart and its function.

301. An oncologist is a doctor who specializes in the medicine, diagnosis, and treatment of tumors.

302. Choose one who specializes in the procedure you'd like to have done and is certified in the specialty by a board recognized by the American Board of Medical Specialties, such as the American Board of Plastic Surgery or the American Board of Facial Plastic and Reconstructive Surgery.

303. Any woman with a congenital heart defect, repaired or not, who is considering pregnancy should talk beforehand with a doctor who specializes in the diagnosis and treatment of heart conditions (cardiologist).

304. After your initial appointment, you may be referred to a doctor who specializes in the diagnosis and treatment of conditions that affect the brain and nervous system (neurologist).

305. Patients should seek a plastic surgeon who specializes in the procedure that she desires to have.

306. This test is performed by a sonographer and then read by a perinatologist, a physician who specializes in the care of pregnant women and their babies.

307. For your personal situation, it is best to follow the advice of your doctor who is preferably a neurologist who specializes in the treatment of multiple sclerosis.

308. Please discuss your specific needs with your doctor, surgeon, or enterostomal therapist (ET), a nurse who specializes in the care of stomas.

309. A physician who specializes in oncology, the study and treatment of cancer, will probably take the lead on treatment.

310. However, Bob Arnold, MD, an internist who specializes in the ethics of doctor-patient relations at the University of Pittsburgh, says sometimes doctors themselves are reluctant to bring up these issues.

311. A naturopathic or holistic doctor, who specializes in the use of herbal remedies to treat medical conditions, can steer you to the right supplements based on your symptoms, the medications you're taking, and your general health.

312. If your doctor suspects you may have a kidney problem, such as nephrotic syndrome, you may be referred to a doctor who specializes in the kidneys (nephrologist).

313. Given your husband's young age, a geriatrician, a doctor who specializes in the elderly, would not be appropriate.

314. A periodontist is a dentist who specializes in the treatment of gum diseases.

315. Oppositional defiant disorder (ODD) is diagnosed and treated by a psychiatrist or psychologist who specializes in the treatment of children.

316. A periodontist is a dentist who specializes in the diagnosis and treatment of periodontal disease.

317. An allergist is a doctor who specializes in the treatment of people who have allergies.

318. After an initial examination, it's likely that the doctor will refer you or your child to a doctor who specializes in the diagnosis and treatment of heart conditions (cardiologist).

319. An evaluation by a heart doctor (cardiologist) or a doctor who specializes in the nervous system (neurologist) may be necessary to look for causes of central sleep apnea.

320. For a second opinion, look to a doctor who specializes in the treatment of your condition.

321. An internist or family practice physician is a medical doctor who specializes in the diagnosis and medical treatment of adults.

322. However, you may then be referred to a doctor who specializes in disorders of blood vessels (vascular specialist) or a doctor who specializes in the heart and circulatory system (cardiologist).

323. If it's thought that you have a liver mass, you may be referred to a doctor who specializes in the digestive system (gastroenterologist) or one who specializes in the liver (hepatologist).

324. However, you may eventually be referred to a rheumatologist a doctor who specializes in the treatment of arthritis and other inflammatory conditions for diagnosis and treatment.

325. If you have any of the problems listed above, the first step toward improving your job outlook is to see a doctor who specializes in the treatment of adult ADHD and get diagnosed so that you can get started on the proper treatment.

326. If you or your doctor thinks you have non-small cell lung cancer, you will need to see an oncologist, which is a doctor who specializes in the treatment of cancer.

327. If your doctor suspects you may have a liver problem, such as nonalcoholic fatty liver disease, you may be referred to a doctor who specializes in the liver (hepatologist).

328. An oncologist is a doctor who specializes in the treatment of cancer.

329. Once your dentist or prosthodontist (a dentist who specializes in the restoration and replacement of teeth) determines what type of appliance is best for you, the general steps are to:

330. Although you're likely to start by seeing your family doctor or general practitioner, you may need to consult an endocrinologist, a doctor who specializes in the hormone-producing (endocrine) glands.

331. Your family physician may refer you to a rheumatologist, a doctor who specializes in the treatment of arthritis and other inflammatory conditions.

332. After going over your history, you may be referred to a urologist or urogynecologist, a physician who specializes in the urinary tract.

333. You can find a therapist who specializes in the anxiety disorders by going to the Anxiety Disorders Association of America (ADAA) website.

334. A neurologist, a doctor who specializes in the brain and nervous system, is best able to diagnose and treat epilepsy.

335. Your doctor may refer you to a specialist such as a doctor who specializes in male genital problems (urologist), a doctor who specializes in the hormonal systems (endocrinologist), a doctor who diagnoses and treats mental health problems (psychiatrist), or another type of specialist.

336. Depending on your particular health concerns, you may go directly to a specialist such as a doctor who specializes in male genital problems (urologist) or a doctor who specializes in the hormonal systems (endocrinologist).

337. He or she may refer you to a doctor who specializes in the diagnosis and treatment of skin conditions (dermatologist).

338. If you or any of your immediate family members (parents or siblings) have hemophilia or are carriers and you are thinking about having a child, you may want to talk to a health professional who specializes in the study of inherited disorders (medical geneticist) before becoming pregnant.

339. If your doctor suspects you may have pancreatitis, you may be referred to a doctor who specializes in the digestive system (gastroenterologist).

340. A psychiatrist is a medical doctor who specializes in the treatment of mood disorders and emotional distress.

341. A gynecologist is a medical doctor who specializes in the areas of women's general and reproductive health, pregnancy, and labor and childbirth.

342. That's a doctor who specializes in the diagnosis and treatment of allergies.

343. The recorder monitors your hearts electrical activity and is inserted at the hospital by an electrophysiologist, a physician who specializes in the electrical activity of the heart.

344. Lithium Aspartate is commonly used in the natural treatment of hyperthyroidism and other thyroid diseases because it helps in the spreading of iodine evenly throughout the body.

345. As a part of methionine and cysteine, it helps in the metabolism of homocysteine.

346. Vitamin B12 is an especially important vitamin for maintaining healthy nerve cells, and it helps in the production of DNA.

347. They also claim that it helps in the development of joint collagen and synovial fluids, also good if proven.

348. If we have that 10 or 15 minutes early notification, it helps in the hospital."

349. It's especially important for pregnant women to get enough folic acid because this helps in the development of your baby's spinal cord and nervous system.

350. It is especially important for pregnant women to get enough folic acid because this helps in the development of your baby's spinal cord and nervous system.

351. Pre-hypertension is now known to increase the likelihood of damage to arteries and the heart, brain, and kidneys, so many doctors are now recommending early treatment, though there is no evidence that this helps in the long run.

352. Rosewater is a wonderful skin toner which helps in the natural firming and toning of skin along with hydrating the skin to make it glow with health.

353. Starting from folic acid (a B complex), which helps in the early brain development, these vitamins help in many aspects of metabolism.

354. I am currently on focalin, which helps in the morning but thats all.

355. All types of cabbage are an excellent source of this fat soluble vitamin that helps in the mineralization of our bones.

356. Joint Vibration Analysis is a diagnostic test for TMJ that helps in the diagnosis of a TMJ problem.

357. Some Polyunsaturated Fatty Acid (PUFA) can lower blood pressure due to its linoleic acid content, a raw material of prostaglandin 2 (PG2) that helps in the contraction and relaxation of muscles.

358. If acupuncture helps in this regard, I think it is a positive... Read more

359. If acupuncture helps in this regard, I think it is a positive contribution.

360. If acupuncture helps in this regard, I think it is a positive... Read more As I have stressed in previous blogs, as I emphasize in my book, The Arthritis Handbook: Improve Your Health and Manage the Pain of Osteoarthritis (Diamedica, 2008), and as I tell my patients every day, joints need movement and exercise to stay healthy.

361.

362. A radiologist is a doctor who specializes in the branch of medicine that uses imaging technologies such as X-ray, computerized tomography (CT) scans, and magnetic resonance imaging (MRI) to diagnose and treat diseases.

363. This is a doctor who specializes in the care of people (especially kids) with asthma and allergies.

364. An allergist is a doctor who specializes in the diagnosis and treatment of allergies.

365. If you've had an allergic reaction, it's important to talk to an allergist, a doctor who specializes in the diagnosis and treatment of allergic disease.

366. For testing, you may be referred to a cardiologist a doctor who specializes in the study of the heart and its function.

367. However, in some cases, you may be referred immediately to a doctor who specializes in the body's hormone-secreting glands (endocrinologist).

368. An endocrinologist is a medical doctor who specializes in the treatment of diabetes and other endocrine disorders.

369. The diagnosis and treatment of Graves' disease is often performed by an endocrinologist-a doctor who specializes in the body's hormone-secreting glands.

370. U.S. researchers have found that heart patients who get their defibrillator from a doctor who specializes in the heart's electrical system fare better than those whose devices are implanted by doctors of other specialties.

371. If you're thought to have a peptic ulcer, you may be referred to a doctor who specializes in the digestive system (gastroenterologist).

372. An endodontist is a dentist who specializes in the causes, diagnosis, prevention, and treatment of diseases and injuries of the human dental pulp or the nerve of the tooth.

373. A urologist is a medical doctor who specializes in the diagnosis and treatment of diseases of the urinary tract and genital organs.

374. Evaluation, diagnosis and long-term management of anaphylaxis are complicated, so you'll probably need to see a doctor who specializes in allergies and immunology.

375. However, in some cases when you call to set up an appointment, you may be referred immediately to a doctor who specializes in skin conditions (dermatologist) or one who specializes in allergies (allergist).

376. Your doctor may choose to give you the vaccine made without use of eggs or send you to a physician who specializes in allergies.

377. You may need to see an allergist, a doctor who specializes in allergies to help you figure out what's causing your rash.

378. Your regular doctor may recommend that you see an allergist, or a doctor who specializes in allergies and their symptoms.

379. A physician who specializes in oncology, the study and treatment of cancer, will probably take the lead on treatment.

380. My salvation was a supportive husband and a good psychologist who specializes in oncology.

381. The three that commonly deal with the disease are internists (an internist is a physician that specializes in adult medicine; also both cardiologists and endocrinologists (who are internist that recieve extra training, i.e. specialists, in the heart and hormone disorders) see patient's with the disease.

382. You might want to ask for a referral to a mental health professional that specializes in adult ADHD and is familiar with how the symptoms manifest in adults.

383. If your health insurance covers anything to do with your peepers without forcing you to buy separate vision insurance, you most likely will be able to see either an ophthalmologist (a medical doctor who specializes in eyes) or an optometrist, also known as a doctor of optometry (O.D.).

384. Has she seen an ophthalmogist, a medical doctor who specializes in eyes, to check way she is losing her vision?

385. To be safe, youll want to consult an attorney who specializes in reproductive law if you decide to pursue IVF.

386. My doctor has asked me not to worry about it..I'm not sure what to do I suggest seeing a doctor who specializes in reproductive endocrinology.

387. So, we have an appointment (Feb 6) with a local child psychiatrist who specializes in ADHD and bipolar.

388. You can also find a professional who specializes in ADHD diagnosis through your health plan, your childs teacher or school counselor, other parents of children with ADHD, or nonprofit organizations such as Children and Adults with Attention-Deficit/Hyperactivity Disorder (CHADD).

389. If you are not currently seeing someone who specializes in ADHD, you may want to ask for a referral.

390. I have found a great new Dr that I have found in my area that specializes in nothing but headaches.

391. I even got an Attorney's name that specializes in nothing but 504.

392. Your team may include a physical therapist, occupational therapist, rehabilitation nurse, rehabilitation psychologist, social worker, dietitian, recreation therapist and a doctor who specializes in physical medicine (physiatrist) or spinal cord injuries.

393. Peterson, who specializes in physical medicine and rehabilitation, is a clinical assistant professor at Thomas Jefferson University Hospital in Philadelphia.

394. A number of specialists will be involved in stabilizing the condition, including a doctor who specializes in nervous system disorders (neurologist) and a surgeon who specializes in spinal cord injuries and other nervous system problems (neurosurgeon), among others.

395. Richard Henrys, an orthopedic surgeon in Miami who specializes in spinal surgery, has had two women patients with spinal compression fractures after doing military presses weightlifting.

396. I have taken action and provided my neurologist's info to my ortho, and also, with the encouragement of my physical therapist, referred myself to an ortho who specializes in spinal care, pain management and rehabilitation.

397. The rehabilitation team, which will include a variety of specialists, will be led by a doctor who specializes in spinal cord injury.

398. A radiologist is a doctor who specializes in medical imaging (like x-rays and mammograms).

399. Can you afford to have a consultation with a lawyer in your area who specializes in medical malpractice?

400. For example, someone who specializes in hand injuries might not be the best choice for a person with a spinal cord injury.

401. A course of splinting and therapy by a therapist who specializes in hand therapy is prescribed and supervised by the hand surgeon.

402. If you haven't consulted a neurologist who specializes in MS, I recommend that you seek a 2nd opinion.

403. With the help of a therapist who specializes in PTSD, I filed a claim with the VA and eventually spent 6 weeks in the North Chicago STDU unit to learn to cope with my disorder.

404. My physical medicine doctor is setting me up with a different neurologist up north that specializes in MS and has a good bed side manner.

405. Check around your area for an nuero that specializes in MS and see how it goes.

406. Look for a yoga class that specializes in MS - or "Gentle Yoga" - or get a DVD by Eric Small (the MS Society will loan you one of his from their library FOR FREE for you to try out at home).

407. with a neurologist that specializes in MS at an MS Center so hopefully I will get some clarification on what is happening to me.

408. If I were you I would try to find a Neurologist that specializes in MS and have them evaluate your husband.

409. I am going to a neurologist that specializes in MS at the end of the month to see if I can get some answers.

410. Because burning mouth syndrome is associated with such a wide variety of other medical conditions, your doctor or dentist may refer you to another specialist, such as a skin doctor (dermatologist), a doctor who specializes in ear, nose and throat problems (ENT) or another type of doctor.

411. If you're experiencing a lot of dizziness, you may be referred to a doctor who specializes in ear, nose and throat complaints.

412. Your health professional may also recommend diagnostic tests or refer you to the proper specialist, such as an otolaryngologist who specializes in ear, nose, and throat conditions.

413. After the initial assessment, you'll likely be referred to a doctor who specializes in ear, nose and throat disorders.

414. If you or your child is diagnosed with acute mastoiditis, you may be put in the hospital to receive treatment and care by an otolaryngologist, a doctor who specializes in ear, nose, and throat disorders.

415. A number of specialists will be involved in stabilizing the condition, including a doctor who specializes in nervous system disorders (neurologist) and a surgeon who specializes in spinal cord injuries and other nervous system problems (neurosurgeon), among others.

416. In some cases you may be referred to a doctor who specializes in nervous system disorders (neurologist).

417. However, in some cases when you call to set up an appointment, you may be referred to a neuro-ophthalmologist or a neurologist a doctor who specializes in nervous system disorders, including diseases of the brain, spinal cord, nerves and muscles.

418. However, in some cases when you call to set up an appointment, you may be referred immediately to a neurologist a doctor who specializes in nervous system disorders, including diseases of the brain, spinal cord, nerves and muscles.

419. With the help of a therapist who specializes in PTSD, I filed a claim with the VA and eventually spent 6 weeks in the North Chicago STDU unit to learn to cope with my disorder.

420. With your symptoms, I would try to find one who specializes in PTSD and trauma.

421. Finding a doctor who specializes in Bipolar treatment may be the way to go, given the circumstances.

422. Have you asked for a referral to a good Psychiatrist who specializes in Bipolar Disorder and related illnesses so that a proper diagnosis can be made.

423. Make sure if you do consider surgery, you check out the surgeon and confirm that he is board certified and that he specializes in this surgery and has significant experience.

424. Be certain you ask the doctor if he is an electrophysiologist and if he specializes in this procedure: It can make a difference in the results you get.

425. The lead researcher of the study, who specializes in adult primary care medicine notes that physical activity has been shown to moderate conditions like blood pressure, diabetes, obesity, abnormal lipid levels, helping to reduce overall risk of heart disease.

426. People with attention deficit hyperactivity disorder (ADHD) may consider hiring an organizational coach who specializes in adult ADHD.

427. If you do not get the cooperation that you are looking for consider seeking a consultation from someone who specializes in treatment of mood disorders.

428. I work with a doc (my ob/gyn) who specializes in treatment of fertility issues including PCOS and understands the migraine issue.

429. It is important to find a find a Psychologist who specializes in treatment of PTSD.

430. Please seek some help from a therapist who specializes in treatment of anxiety disorders or check into some clinical trials.

431. A urologist, or doctor who specializes in treatment of conditions of the kidney, bladder and genitals, is usually consulted.

432. It is estimated that a neurologist (a physician who specializes in nerve and brain disorders) sees five patients with PCS per month.

433. Your doctor may refer you to another specialist such as a neurologist, a doctor who specializes in nerve and brain conditions.

434. After that, see someone that specializes in the type of cancer involved, because he or she will have a better handle on the different courses of treatment.

435. Podiatrist: A podiatrist is a physician that specializes in the evaluation and treatment of diseases of the foot.

436. Therapists that specializes in the cognitive behavioral approach can be especially helpful.

437. The findings, which appear in the journal Fertility and Sterility, stem from a study of all mother-infant pairs admitted over a 3-year period to a hospital unit that specializes in the care of mothers with mood disorders or exhaustion, and infants with sleeping or feeding problems.

438. Regardless, patients with acute aortic dissections should be emergently referred to a center that specializes in the treatment of aortic diseases.

439. eventaully about a month before i seeked help for my addiction i had FINALLY gone to see a doctor that specializes in the stomach.

440. When was the last time you saw a Dr. and what did he/she say???? Maybe you need to go to a surgeon that specializes in the issues you are having, I am unsure what type of Dr.'s you have seen in the past.

441. I got there and the Dr. that specializes in the area of behavioral told me that my child was mentally retarded and would never get past where he is today.

442. The dialysis treatment prescription and regimen is usually overseen by a nephrologist (a doctor that specializes in the kidney).

443. a urologist is a doctor that specializes int he urinary tract of males and females and the reproductive organs of males.

444. Just in case, i really recommend you seeing a physician or any doctor that specializes in the Breast Society.

445. I feel...if you have a bad tooth, you go to a dentist, if you have cancer, you go to an oncologist....If you have a particular disease, you go to the doctor that specializes in the care and treatment of that disease.

446. In July 2000, a mobile tower crane fell into an adjacent water tower, causing the death of a 29- year-old man working for a tank company that specializes in the relocation of used water towers.

447. The first thing you should know when you go to an appointment to diagnose oral cancer is that you will probably be referred to a dentist or periodontist, a dentist that specializes in the gums.

448. It sounds to me like you need to go see a pain management Doctor and a Doctor that specializes in the spine, it could be a Neurosurgeon or an Orthopedic Doctor that specializes in spines only.

449. If you are looking for a facility for a loved one, who suffers from dementia, then it is important that you consider a community that specializes in the type of care they need.

450. A baby with TGA needs to be treated in a hospital that specializes in the treatment of children with congenital heart disease.

451. Your doctor may refer you to a doctor who specializes in blood diseases (hematologist).

452. If you're diagnosed with a type of anemia that requires more complex treatment, such as aplastic anemia or anemia caused by other diseases, you may be referred to a doctor who specializes in blood disorders (hematologist).

453. If your doctor suspects you have a myelodysplastic syndrome, you may be referred you to a doctor who specializes in blood disorders (hematologist).

454. However, in some cases, you may be referred immediately to a doctor who specializes in blood cell diseases (hematologist).

455. Patients should seek care from a doctor who specializes in blood disorders (hematologist) or a clinic that is experienced in treating sickle cell disease.

456. Once sickle cell anemia is diagnosed, you'll likely be referred to a doctor who specializes in blood disorders (hematologist) or a pediatric hematologist.

457. In more serious cases of PAD, where leg pain is so severe and persistent that walking is problematic, the lifestyle changes and medications may be combined with a minimally invasive procedure performed by a physician who specializes in blood vessel health, or through surgery to create a detour around the vessel blockages to restore blood flow.

458. If you're found to have monoclonal gammopathy of undetermined significance, you may be referred to a hematologist, a doctor who specializes in blood disorders.

459. You may then be referred to a doctor who specializes in blood disorders (hematologist).

460. You may be referred to a doctor who specializes in blood diseases (hematologist).

461. However, because porphyria can be difficult to diagnose, when you call to set up an appointment, you may be referred immediately to a doctor who specializes in blood disorders (hematologist).

462. However, you may then be referred to a doctor who specializes in blood disorders (hematologist).

463. You are also likely to be referred to a doctor who specializes in blood disorders (hematologist) for further evaluation and treatment.

464. Find a physical therapist, exercise physiologist, or chiropractor who specializes in back care.

465. I'm afraid I don't know the answers to your other questions about sciatica, but I will bring your question to the attention of Dr. Lasich, who specializes in back problems.

466. Find a medical doctor who specializes in back injuries.

467. However, your child will need lifelong follow-up care with a heart doctor (cardiologist) who specializes in congenital heart disease.

468. Following are some of the procedures babies with tricuspid atresia may require: Follow-up care To monitor his or her heart health, your baby will need lifelong follow-up care with a cardiologist who specializes in congenital heart disease.

469. After surgery After corrective surgery, your baby will need lifelong follow-up care with a heart doctor (cardiologist) who specializes in congenital heart disease to monitor his or her heart health.

470. Follow-up care After surgery or a transplant, your baby will need lifelong follow-up care with a heart doctor (cardiologist) who specializes in congenital heart disease, to monitor his or her heart health.

471. I wondered if anyone knows any doctors in Utah, around the Salt Lake City or Ogden area that specializes in hormone therapy?

472. You may want to look for a doctor that specializes in hormone imbalances to see if this is an issue.

473. I'm so angry about this but I feel captive because it's so hard to find a neuro that specializes in migraines.

474. My neurologist that specializes in migraines recently put me on Seroquel and the results have been dramatic.

475. I found a good neurologist that specializes in migraines.

476. thanks michel You should find you your own Dr. that maybe a Neurologist, maybe even one that specializes in migraines.

477. I would definately seek a 2nd opinion and be sure it would a specialist that specializes in migraines.

478. How do we find a neurologist that specializes in migraines in our area?

479. My brother sugested that I contact this web site and get ahold of a Dr. in CT that specializes in migraines.

480. She also said that I needed inpatient hospitialization that specializes in migraines for 2wks.

481. On the following day, my colleague, Dr.Evan Nadler, a pediatric surgeon who specializes in obesity and bariatric surgery, delivered a talk on metabolic abnormalities associated with obesity and treatment to an international audience of pediatric specialists at the Arab Health Summit in Dubai.

482. Even after a surgeon, mental health professional, dietician and physician who specializes in obesity has determined the patient should have bariatric surgery, the patient is told to try six months or a year of additional, medically supervised weight loss attempts before trying surgery, Shikora said.

483. "Whole-grain cereal is a great replacement for high-fat breakfast food or as a replacement for no breakfast at all, since breakfast is the most important meal of the day," said To, who specializes in obesity and diabetes management.

484. What to do: Report the rashes to your regular doctor or a doctor who specializes in skin disorders to evaluate and rule out other causes.

485. First your doctor or dermatologist (a doctor who specializes in skin problems) will try to determine the underlying cause of your hair loss.

486. However, he or she may then refer you to a doctor who specializes in skin disorders (dermatologist).

487. After diagnosis, though, it is often recommended that people with rosacea see a dermatologist-a doctor who specializes in skin conditions.

488. A dermatologist is a physician who specializes in skin diseases, specifically in treating the medical, surgical, and cosmetic conditions of the hair, skin, and nails.

489. He or she may suggest treatment, or may refer you to a doctor who specializes in skin disorders (dermatologist) or circulatory disorders (cardiologist).

490. Your family doctor may refer you to a dermatologist, a doctor who specializes in skin disorders.

491. However, in some cases when you call to set up an appointment, you may be referred immediately to a doctor who specializes in skin conditions (dermatologist) or one who specializes in allergies (allergist).

492. A dermatologist is a doctor who specializes in skin and hair conditions.

493. He or she may refer you to a doctor who specializes in skin disorders (dermatologist).

494. Depending on your situation and the outcome of any tests, you may be referred to a doctor who specializes in skin diseases (dermatologist) or to a doctor who specializes in cancer treatment (oncologist).

495. Depending on your childs symptoms, you may want to seek treatment from various specialists such as a rheumatologist who specializes in joints and connective tissue diseases, a nephrologist who specializes in kidney diseases, a dermatologist who specializes in skin health, or infectious diseases specialist.

496. If you and your doctor aren't able to find a medication that works well for you, he or she may recommend that you see a doctor who specializes in skin disorders (dermatologist).

497. He or she may then refer you to a doctor who specializes in skin disorders (dermatologist) for diagnosis and treatment.

498. I'd get my second opinion from a university hospital based dermatologist if possible, or at minimum make sure I am being seen by a dermatologist who specializes in skin cancers and is board certified.

499. However, he or she might refer you to a doctor who specializes in skin disorders (dermatologist).

500. If your warts don't respond to conservative treatments, you might be referred to a doctor who specializes in skin disorders (dermatologist).

501. Cellulitis may be diagnosed and treated by a family doctor, an infectious disease specialist, a doctor who specializes in skin diseases (dermatologist), or in the case of orbital cellulitis, an eye doctor (ophthalmologist).

502. You might be referred to a doctor who specializes in skin conditions (dermatologist).

503. He or she may refer you to a doctor who specializes in skin disorders (dermatologist) or hormone problems (endocrinologist).

504. However, in some cases when you call to set up an appointment, you may be referred immediately to either a doctor who specializes in skin conditions (dermatologist) or one who specializes in foot conditions (podiatrist).

505. If your rash is more severe, you may want to see your primary care doctor or a doctor who specializes in skin disorders (dermatologist) to be sure it's heat rash and not another skin disorder.

506. You may then be referred to a doctor who specializes in skin disorders (dermatologist).

507. Geriatrician is a doctor who specializes in geriatric medicinea practice that is focused on helping to prevent and treat health problems in elderly people.

508. I would get a qualified psychaitrist who specializes in geriatric or older patients [forgive me if 58 is too young for a geriatric doctor, though].

509. He or she may refer you to a doctor or surgeon who specializes in foot disorders.

510. However, in some cases when you call to set up an appointment, you may be referred immediately to either a doctor who specializes in skin conditions (dermatologist) or one who specializes in foot conditions (podiatrist).

511. If not, I would suggest contacting your doctor or a psychologist that specializes in anxiety.

512. Once this is done, they should refer you to a therapist or doctor in your area that specializes in anxiety in children.

513. In particular, see if you can find one that specializes in anxiety disorders.

514. A doctor that specializes in anxiety once said that you can't be anxious while you are breathing deeply.

515. Exposure therapy is best accomplished with an experienced therapist that specializes in anxiety disorders.

516. I do see a Dr. that specializes in bone and mineral research, so she doesn't treat many other disorders than bone like Endo's, Rheumie's and Internist's do.

517. Does your great grandson have a Dr that specializes in bone disorders?

518. If your doctor suspects you may have a form of peripheral neuropathy, he or she may refer you to a neurologist, a doctor who specializes in diseases of the nerves.

519. Although your symptoms may prompt you to visit your family doctor or a general practitioner, you'll likely be referred to a doctor who specializes in diseases of the digestive system (gastroenterologist) to diagnose and treat Zollinger-Ellison syndrome.

520. If your doctor suspects you have non-Hodgkin's lymphoma, you may be referred to a doctor who specializes in diseases that affect the blood cells (hematologist).

521. Here is the difference: an ophthalmologist is a medical doctor who specializes in diseases of the eye.

522. If your doctor thinks you may have a systemic bleeding disorder, you will probably need to see a hematologist, a doctor who specializes in diseases of the blood.

523. However, once diagnosed, or in order to confirm diagnosis, you will most likely be referred to a rheumatologist, who specializes in diseases that affect the muscles and joints.

524. If your doctor determines you may have chronic lymphocytic leukemia, you may be referred to a doctor who specializes in diseases of the blood and bone marrow (hematologist).

525. If your doctor or dentist feels you may have mouth cancer, you may be referred to a dentist who specializes in diseases of the gums and related tissue in the mouth (periodontist) or to a doctor who specializes in diseases that affect the ears, nose and throat (otolaryngologist).

526. If the results of the DRE and PSA tests are indicative of a significant prostate disorder, the examining physician usually refers the patient to a urologist, a physician who specializes in diseases of the urinary tract and male reproductive system.

527. If your doctor suspects you may have liver cancer, you may be referred to a doctor who specializes in diseases of the liver (hepatologist) or to a doctor who specializes in treating cancer (oncologist).

528. If your doctor suspects you may have cancer or another disease that affects your throat, you may be referred to a doctor who specializes in diseases and conditions that affect the ears, nose or throat (otolaryngologist, or ENT specialist).

529. During a biopsy, a urologist (a doctor who specializes in diseases of urinary and sex organs in men, and urinary organs in women) removes tissue samples, usually with a needle.

530. You also may be referred to a doctor who specializes in diseases of the joints, bones and muscles (rheumatologist).

531. If your doctor suspects you may have proctitis, you may be referred to a doctor who specializes in diseases of the digestive system (gastroenterologist).

532. Your doctor may refer you to a hepatologist, a doctor who specializes in diseases of the liver.

533. If your doctor suspects you may have a thyroid problem, you may be referred to a doctor who specializes in diseases of the endocrine system (endocrinologist).

534. However, you may then be referred to a doctor who specializes in metabolic disorders (endocrinologist) or a doctor who specializes in diseases of the joints, muscles or bones (rheumatologist).

535. Depending on the suspected cause, your doctor may refer you to a doctor who specializes in treating ear, nose and throat disorders (otorhinolaryngolist), a doctor who specializes in treating digestive disorders (gastroenterologist), or a doctor who specializes in diseases of the nervous system (neurologist).

536. Depending on the findings of the examination, your doctor may refer you to a urologist (a doctor who specializes in diseases of the urinary tract) or neurologist (a doctor who specializes in diagnosing and treating diseases of the nervous system).

537. HPV treatment as well as cervical cancer screening and diagnosis are usually handled by a gynecologist, a physician who specializes in diseases and conditions of the female reproductive system.

538. People with symptoms like these may see their family doctor or a urologist, a doctor who specializes in diseases of the urinary system.

539. "We all shed around one hundred to one hundred fifty hairs per day," says Paradi Mirmirani, MD, a dermatologist in Vallejo, CA, who specializes in hair disorders.

540. If their edges aren't lined up perfectly, then hair is dull, brittle, and has no shine," says Hanjani Galant, who specializes in hair and hair disorders.

541. "On average, we lose fifty to a hundred hairs a day," says Francesca Fusco, MD, a New York City dermatologist who specializes in hair loss.

542. I would suggest you speak with a dermatologist who specializes in hair loss.

543. However, in some cases when you call to set up an appointment you may be referred immediately to a doctor who specializes in conditions affecting the female reproductive tract (gynecologist), one who specializes in hormonal disorders (endocrinologist) or one who specializes in both areas (reproductive endocrinologist).

544. Choose a therapist who specializes in both individual and couples counseling.

545. If your child has an illness or injury that requires emergency department visit, it is best to choose a hospital that specializes in children.

546. I would suggest contacting a mental health provider that specializes in children and request a complete evaluation.

547. Most people are denied and have to apply again, sometimes they have to hire an Attorney who specializes in disability cases.

548. Working with a lawyer who specializes in disability benefits could make this process easier.

549. Call your State (not county) bar association and tell them you need an attorney who specializes in disability claims, that you feel you have been wrongfully fired on the basis of a disability.

550. However, in some cases when you call to set up an appointment you may be referred immediately to a doctor who specializes in conditions affecting the female reproductive tract (gynecologist), one who specializes in hormonal disorders (endocrinologist) or one who specializes in both areas (reproductive endocrinologist).

551. A doctor who specializes in hormonal disorders (endocrinologist) generally coordinates diabetes care, but your health care team likely will include: Once your blood sugar is under control, your endocrinologist likely will recommend checkups every few months.

552. Since the information obtained from a genetic test can have a profound impact on your life, you may want to see a doctor who specializes in genetics (geneticist) or a genetic counselor.

553. Since the information obtained from karyotyping can have a profound impact on your life, you may want to see a doctor who specializes in genetics (geneticist) or a genetic counselor.

554. If you can't find one in your area, try one who specializes in sports medicine.

555. He is a board certified surgeon who specializes in sports medicine and diabetic foot reconstruction.

556. You may initially bring your signs and symptoms to the attention of your family physician, but he or she may refer you to a doctor who specializes in sports medicine or rheumatology the treatment of conditions that affect the joints.

557. If these people cannot help, then a doctor who specializes in sports medicine may be able to suggest a competent sports professional who can guide you.

558. Try finding a doctor who specializes in sports medicine or physical medicine.

559. A urologist who specializes in incontinence can guide you to pelvic floor exercises and medications, or possibly a surgical fix.

560. Ask your doctor for a referral to a urologist who specializes in incontinence.

561. I was put on a low dose of wellbutrin 75 mg because I am depressed and also just made an apt. to see a pyschiatrist who specializes in patients who have chronic pain.

562. I'm under the care of a neurologist in Grosse Pointe Woods, Mi., who specializes in patients with migraine.

563. If you're diagnosed with endometrial cancer, you're likely to be referred to a doctor who specializes in cancers of the female reproductive system (gynecologic oncologist).

564. If it's determined that you have vaginal cancer, you'll likely be referred to a doctor who specializes in cancers of the female reproductive system (gynecologic oncologist).

565. If your doctor suspects you may have a stomach problem, you may be referred to a doctor who specializes in gastrointestinal diseases (gastroenterologist). Once stomach cancer is diagnosed you may be referred to a cancer specialist (oncologist) or a surgeon who specializes in operating on the digestive tract.

566. A vagotomy is usually performed by a board-certified surgeon, either a general surgeon who specializes in gastrointestinal surgery or a gastrointestinal endoscopic surgeon.

567. Is there any other diagnosis available or medical group that specializes in these disorders?

568. case to, which we used tying in the migraines and some hi stress physical conditions together, i'm not sure about the ss clock time, but if your using a lawyer that specializes in these cases, he's got it all down pat and you need him.

569. I'm not sure HealthCentral has a site that specializes in these types of issues, which I guess is why you're on the rheumatoid arthritis site?

570. It is so very important to choose a surgeon that specializes in these ops only, even if u have to travel out of state.

571. The nurse on the phone suggests that we have my sons pediatrician make a referral to get him seen by a second neurologist who specializes in seizures.

572. I know even before I ask you're going to tell me I need to ask a neurologist- one who specializes in seizures.

573. While there are many cold medicines and treatments that soothe miserable cold symptoms, there is nothing that cures a cold.

574. Of course, in today's common usage, the term antibiotic is used to refer to almost any drug that cures a bacterial infection.

575. There are no cures for mitochondrial diseases, only treatments and therapies to help patients and caregivers manage the symptoms.

576. There are no cures for mitochondrial diseases, only treatments and therapies, which can help to lessen symptoms and slow the disease.

577. There are no cures for viral infections, due in part to the difficulty of developing drugs that adversely affect only the virus and not the host.

578. There are no cures for viral skin infections, and such infections can reoccur.

579. Additionally, the prostate surrounds the male urethra, which is the passageway for urine to flow from the... Read more In the 21st century, medicine has seen unrivaled advances, enabling patient care that includes finding cures for previously incurable diseases.

580. In the 21st century, medicine has seen unrivaled advances, enabling patient care that includes finding cures for previously incurable diseases.

581. Comment: Taking Actos for 10 years, and 6 months ago, I began having symptoms of a heart attack.

582. I've had frequent UTI's since I lost my virginity five years ago, in March I began having symptoms of a UTI (which is virtually second nature to me) and went along with the usual antibiotics.

583. However, stop taking meclofenamate and immediately seek medical attention if you notice any of the following symptoms of a serious allergic reaction: rash/blisters, itching/swelling (especially of the face/tongue/throat), severe dizziness, trouble breathing.

584. Seek immediate medical attention if you notice any of the following symptoms of a serious allergic reaction: new fever, rash, itching/swelling (especially of the face/tongue/throat), severe dizziness, trouble breathing.

585. However, seek immediate medical attention if you notice any of the following symptoms of a serious allergic reaction: rash, skin lesions/sores, itching/swelling (especially of the face/tongue/throat), severe dizziness, trouble breathing, new or worsening swelling/pain in the joints, swollen glands, loss of appetite/weight loss, chest pain, fast/irregular heartbeat, severe stomach/abdominal pain, yellowing eyes/skin, dark urine.

586. However, stop taking mefenamic acid and immediately seek medical attention if you notice any of the following symptoms of a serious allergic reaction: rash/blisters, itching/swelling (especially of the face/tongue/throat), severe dizziness, trouble breathing.

587. However, stop taking tolmetin and immediately seek medical attention if you notice any of the following symptoms of a serious allergic reaction: rash/blisters, itching/swelling (especially of the face/tongue/throat), severe dizziness, trouble breathing.

588. However, get medical help right away if you notice any of the following symptoms of a serious allergic reaction, including: rash, skin lesions/sores, itching/swelling (especially of the face/tongue/throat), severe dizziness, trouble breathing, new or worsening swelling/pain in the joints, swollen glands, chest pain, loss of appetite/weight loss, fast/irregular heartbeat, severe stomach/abdominal pain, yellowing eyes/skin, dark urine.

589. Seek immediate medical attention if you notice any of the following symptoms of a serious allergic reaction:

590. However, get medical help right away if you notice any of the following symptoms of a serious allergic reaction: fast/irregular heartbeat, flushing of the face, rash, itching/swelling (especially of the face/tongue/throat), severe dizziness, trouble breathing, fever/chills.

591. If you have any symptoms of these, callor other emergency services.Symptoms of a heart attack include: Severe chest pain, also described as discomfort, pressure, squeezing, or heaviness.

592. If you have any symptoms of these, call 911 or other emergency services.

593. Seek immediate medical attention if you have any symptoms of these side effects, including weakness on one side of the body, slurred speech, vision changes, severe stomach/abdominal pain, dark urine, persistent nausea/vomiting, yellowing eyes/skin, mental/mood changes (such as confusion), severe/persistent headache.

594. APR patients were those who experienced symptoms of nasal hypersecretion and nasal congestion or sneezing when exposed to specific perennial allergens (e.g., dust mites, molds) and were skin test positive to these allergens.

595. NAPR patients were those who experienced symptoms of nasal hypersecretion and nasal congestion or sneezing throughout the year, but were skin test negative to common perennial allergens.

596. APR patients were those who experienced symptoms of nasal hypersecretion and nasal congestion or sneezing when exposed to specific perennial allergens (e.g., dust mites, molds) and were skin test positive to these allergens.

597. NAPR patients were those who experienced symptoms of nasal hypersecretion and nasal congestion or sneezing throughout the year, but were skin test negative to common perennial allergens.

598. When you begin taking this medicine, you or your child will be given a warning card which describes symptoms of severe allergic reactions that may be caused by abacavir.

599. When you or your child begin taking this medicine, you will be given a warning card which describes symptoms of severe allergic reactions that may be caused by abacavir, lamivudine, and zidovudine combination.

600. When you begin taking this medicine, you will be given a warning card which describes symptoms of severe allergic reactions that may be caused by abacavir and lamivudine combination.

601. If you develop symptoms of the disease, be sure to see a doctor.

602. If you develop symptoms of the disease, see your doctor for a proper diagnosis.

603. However, if you develop symptoms of the flu such as fever, cough, or sore throat, you should ask someone who is not sick to care for your baby.

604. If you have symptoms of meningitis or encephalitis, a lumbar puncture, also called a spinal tap, may be done to look for antibodies and signs of infection in the cerebral spinal fluid, which surrounds the brain and spinal cord.

605. If you have symptoms of meningitis, seek help immediately.

606. Contact your doctor immediately if you notice any symptoms of dehydration, such as fast heartbeat or dizziness/lightheadedness.

607. Contact your doctor promptly if you notice any symptoms of dehydration, such as unusual decreased urination, unusual dry mouth/increased thirst, lack of tears, dizziness/lightheadedness, muscle weakness/cramping, or pale/wrinkled skin.

608. Contact your doctor promptly if you notice any symptoms of dehydration such as unusual decreased urination, unusual dry mouth/increased thirst, lack of tears, dizziness/lightheadedness, or pale/wrinkled skin.

609. Strong evidence suggests that MS is caused by the immune system causing inflammation and attacking the myelin, which is the coating surrounding the nerve and nerve fibers.

610. Some people believe that MS is caused by the mercury in dental fillings, but there is no scientific evidence to support this.

611. Because rickets is caused by a vitamin D deficiency that leads to weak bones, it affects children differently than adults.

612. If rickets is caused by a metabolic problem, a prescription for vitamin D supplements may be needed.

613. Researchers have found that dyslexia is caused by a difference in the way the dyslexic brain processes information.

614. Because dyslexia is caused by a difference in the structure and function of specific areas of the brain, there is no cure.

615. Crepitus can be broadly grouped into symptoms associated with crepitus in joints, as well as symptoms associated with crepitus due to air in soft tissues.... Read more about crepitus symptoms Crepitus is caused by tissues rubbing together in an abnormal way.

616. Crepitus is caused by tissues rubbing together in an abnormal way.

617. Most commonly, it is caused by changes in the inner ear that occur as you grow older.

618. A more current model suggests that when the visual aura is present, it is caused by changes in blood flow patterns in the brain and cortical spreading depression (CSD), which refers to decreased activity on the surface of the brain.

619. Unlike fear, which is caused by realistic, known dangers, anxiety can be more difficult to identify and to alleviate.

620. Unlike fear, which is caused by realistic, known dangers, anxiety can be more difficult to identify and alleviate.

621. Even though tuberculosis (TB) can be caused by three different bacteria, the term almost always refers to the contagious, potentially fatal infection that is caused by the bacterium Mycobacterium tuberculosis.

622. Glanders is an infectious disease that is caused by the bacterium Burkholderia mallei.

623. Shingles is a painful localized skin rash often with blisters that is caused by the varicella zoster virus (VZV), the same virus that causes chickenpox.

624. Cholera is a bacterial disease that is caused by the activity of Vibrio cholerae in the intestine, which results in severe diarrhea.

625. Orchitis is inflammation in the testes that is caused by the mumps virus or bacteria.

626. Borborygmus: A gurgling, rumbling, or squeaking noise from the abdomen that is caused by the movement of gas through the bowels.

627. Even though the condition may be short term, it is important to relieve the itch and pain that is caused by the skin rashes.

628. A food allergy is an abnormal response of the immune system that is caused by the protein in certain foods.

629. But as the terms periodontal disease, gingivitis, and periodontitis are most commonly used, they refer to disease that is caused by the buildup of dental plaque.

630. "If a child is not vaccinated, he or she will come down with chicken pox if they come into contact with an adult who has shingles," he said, referring to a painful adult condition that is caused by the same virus.

631. Blastomycosis is a rare infectious multisystem disease that is caused by the fungus Blastomyces dermatitidis.

632. I don't know if it is something that is caused by the spinal tap or if it is just a new symptom.

633. There is a small subset of patients with urinary incontinence that is caused by the urethra always remaining "open".

634. It is my opinion that schizophrenia is NOT an actual disease, however it is a symptom of poor oxygen circulation to the brain that is caused by the lungs.

635. Progressive multifocal leukoencephalopathy (PML) is a disease that is caused by the reactivation of a common virus in the central nervous system of immune-compromised individuals.

636. Sydenham chorea is a type of chorea that is caused by the streptococcal bacteria.

637. Diabetes is a chronic disease that is caused by the bodys inability to use glucose (blood sugar) properly due to a lack of or defects in insulin production.

638. Gonorrhea is a highly contagious sexually transmitted disease that is caused by the bacterium Neisseria gonorrhoeae.

639. Meningococcal disease can refer to any illness that is caused by the type of bacteria called Neisseria meningitidis, also known as meningococcus [muh-ning-goh-KOK-us].

640. Ribavirin for inhalation is used to treat severe pneumonia in infants and young children that is caused by the respiratory syncytial virus (RSV).

641. Trisomy 18, also called Edward's syndrome, is a genetic disease that is caused by the presence of an additional copy (or part of an additional copy) of chromosome 18.

642. IatrogenicA condition that is caused by the diagnostic procedures or treatments administered by medical professionals.

643. Mumps is a contagious disease that is caused by the mumps virus.

644. Leprosy is caused by the organism Mycobacteriumleprae.

645. Leprosy is caused by the bacterium Mycobacterium leprae, which attacks the peripheral nerves of affected people.

646. Leprosy is caused by the bacteria Mycobacterium leprae.

647. Arteriolosclerosis is caused by damage to the inner and middle layers of small blood vessels as the result of diabetes and high blood pressure.

648. Arteriolosclerosis is caused by damage to the inner walls of the body's blood vessels as the result of certain health conditions, such as diabetes and high blood pressure.

649. Often it's just that simple, but sometimes it is caused by something more serious, like diabetes.

650. Even if it is caused by something physical, erectile dysfunction can create stress and relationship tension.

651. If bronchitis is caused by a bacterial infection and doesn't get better on its own, an antibiotic may be prescribed.

652. When bronchitis is caused by a virus or irritation in the air (like cigarette smoke) , antibiotic treatment will not help it get better.

653. Diphtheria is caused by a bacterium called Corynebacterium diphtheria.

654. Diphtheria is caused by a bacterium that reproduces on the lining of the nose, throat, and trachea.

655. As you age, your chance of developing osteoarthritis, which is caused by wear and tear, increases.

656. An estimated 21 million adults in the United States have osteoarthritis, which is caused by wear and tear on cartilage in the joints.

657. When symptoms do appear, they generally occur one to three weeks after exposure to the infection.... Read more about chlamydia symptoms Chlamydia is caused by a bacterial infection of the genital tract by the bacterium Chlamydia trachomatis.

658. Chlamydia is caused by a bacterial infection of the genital tract by the bacterium Chlamydia trachomatis.

659. Dengue is caused by any one of four related viruses transmitted by mosquitoes.

660. Dengue is caused by any of four dengue viruses, which are transmitted by Aedesmosquitoes.

661. TB is a disease that is caused by bacteria which are spread from person to person through the air.

662. Chlamydia is an STI that is caused by bacteria called Chlamydia trachomatis.

663. Much less commonly it may be the result of an injury from an instrument such as a urinary catheter or exposure to an irritating chemical such as an antiseptic or a spermicide Gonococcal urethritis, commonly called clap, is a sexually transmitted disease that is caused by bacteria called Neisseria gonorrhoeae.

664. Doctors use antibiotics to treat pneumonia that is caused by bacteria.

665. Gonorrhea is a sexually transmitted disease (STD) that is caused by bacteria called Neisseria gonorrhoeae.

666. Tetanus is a sometimes fatal disease that is caused by bacteria.

667. Typhoid fever is a contagious disease that is caused by bacteria called Salmonella typhi.

668. Since Osteomyelitis is an infection that is caused by bacteria, there is nothing else that will cure it or control it besides antibiotics!! Depending on which bug is causing your Osteo, that will determine which antibiotic you'll need to go on.

669. A lot of acne is caused by bacteria and oil of oregano is a very optent killer of bacteria.

670. Since acne is caused by bacteria that invade oil trapped inside clogged pores, washing your face gently with a mild cleanser twice a day can help keep your skin healthy.

671. There may be other medical causes for the symptom of an over-reactive gag reflex, so you want to allow your doctor to investigate first, before assuming that it is caused by anxiety.

672. If it is caused by anxiety, there are treatments for this as well.

673. Roseola is caused by two forms of the herpes virus.

674. Roseola is caused by two common viruses.

675. Atherosclerosis is caused by the accumulation in the bloodstream of fat, cholesterol, and other substances that build up on the walls of arteries and form hard structures called plaques.

676. However, symptoms of moderate to severe atherosclerosis depend on which arteries are affected.... Read more about atherosclerosis symptoms Atherosclerosis is caused by the accumulation in the bloodstream of fat, cholesterol, and other substances that build up on the walls of arteries and form hard structures called plaques.

677. Exposure to the fumes of zinc chloride may result in a severe pneumonitis that is caused by irritation of the respiratory tract (Gafafer 1964/Ex.

678. Occipital neuralgia (ON), a chronic pain disorder that is caused by irritation or injury to the occipital nerve, or area around the nerve which is located in the back of the scalp.

679. Anthrax is caused by the bacteriumBacillus anthracis, which produces spores that spread the infection.

680. Anthrax is caused by the bacteria Bacillus anthracis.

681. Sometimes pharyngitis is caused by a bacterial infection or tonsillitis.

682. Less commonly, pharyngitis is caused by a bacterial infection.

683. Acromegaly is caused by the pituitary gland overproducing growth hormone (GH) over time.

684. Acromegaly is caused by the prolonged overproduction of growth hormone (GH) by the pituitary gland.

685. Although the exact cause of MS is unknown, researchers believe that MS is caused by a person's immune system attacking myelin in the brain and spinal cord.

686. Current data suggests that MS is caused by a combination of environmental and genetic factors, with some studies indicating that hormones and certain viruses (like mumps, herpes, and chicken pox) may play a role.

687. Actually, for the past one hundred years, scientists have suspected that MS is caused by a virus attack.

688. As the theory goes, MS is caused by a lack of blood flow to the brain.

689. Besides lack of recognition, it is also possible that MS is caused by a "screw-up" in the regulation of the immune system - that a helper-inducer subset of immune system cells, which help keep the immune response going, is numerically or functionally overrepresented.

690. Current data suggests that MS is caused by a combination of environmental and genetic factors, with some studies indicating that hormones and certain viruses may also play a role.

691. Autoimmune thyroiditis (AT), also known as Hashimoto's disease, is a chronic inflammatory disorder of the thyroid gland that is caused by abnormal blood antibodies and white blood cells that mistakenly attack and damage healthy thyroid cells.

692. Polycystic ovarian syndrome is a hormonal condition that is caused by abnormal ovarian function.

693. Loss of taste is caused by interruption of the transfer of taste sensations to the brain, or by a problem with the way the brain interprets these sensations.

694. Loss of taste is caused by interruption of the tran... Read more about loss of taste introduction Inflammation and infection of the upper respiratory tract, sinuses, mouth, and tongue can result in loss of taste.

695. This is caused by the slightly elevated levels of bilirubin in your blood.

696. This is caused by the thinning (atrophy) of vaginal tissues, especially those surrounding the opening of the vagina, due to a lack of estrogen hormones.

697. This is caused by the sudden extension -backward movement of the neck- and flexion - forward movement of the neck.

698. This is caused by the kneecap sliding up and down over the joint but slightly off center.

699. This is caused by the acidity of the food or beverage you eat combined with the acid that the bacteria in all our mouths produce.

700. This is caused by the anxiety you may feel at the doctor's office or in a hospital.

701. This is caused by the response of the immune system.

702. This is caused by the nerves of the bladder firing signals to the brain that the bladder needs to empty far too often.

703. This is caused by the lack of progesterone which is important to the quality of our sleep.

704. This is caused by the pressure of air on the water of the fixture being greater than the pressure of air in the waste pipe.

705. This is caused by the unlodging of the mucosal plug that blocks the entrance of your cervix to prevent infection.

706. IBC is caused by cancer cells that have gotten into the lymph vessels in your skin, and are blocking fluid drainage; thus the redness and possible swelling.

707. The redness, dimpling, pain, and swelling from IBC is caused by cancer cells clogging the lymph vessels of the skin of the breast.

708. Ringworm is caused by a fungus, or yeast, not a worm.

709. Ringworm is caused by a fungus that grows on the skin.

710. It is caused by damage to parts of the brain that control movement.

711. It is caused by damage to the nerves that control the pace at which food leaves the stomach and gets processed in the gut.

712. It is caused by damage to the vagus nerve, which regulates the digestive system.

713. It is caused by damage to the parts of the brain that are involved in speaking, and involves the loss or impairment of existing speech abilities.

714. Severe acute respiratory syndrome (SARS) is a recently recognized febrile severe lower respiratory illness that is caused by infection with a novel coronavirus, SARS-associated coronavirus (SARS-CoV).

715. Colitis that is caused by infection usually goes away when treated with medicine.

716. Malaria is caused by one of several species of the Plasmodium parasite.

717. Malaria is caused by one of four protozoan species of the genus Plasmodium: P. falciparum, P. vivax, P. ovale, and P. malariae and is transmitted by the bite of an infected female Anopheles mosquito.

718. Malaria is caused by one of four kinds of protozoan parasites.

719. Between 1967 and 1972, an epidemiologic study was conducted in an East German chemical plant to assess the risk of cancer associated with exposure to aldol and aliphatic aldehydes produced from the dimerization of acetaldehyde [Bittersohl 1975].

720. 1-117) rat kidney data and multistage model is a reasonable approach for estimating the risk of cancer associated with exposure to chloroform.

721. In contrast, disease associated with 2009 H1N1 influenza is continuing to increase in southern Africa and more Africa countries have reported their first cases.

722. In contrast, disease associated with 2009 H1N1 influenza is continuing to increase in southern Africa, and more African countries have reported their first cases.

723. Spend some time understanding the side effects and risks associated with the procedure.

724. The purpose of this activity is to facilitate research on risks associated with the changing organization of work by broadly disseminating methodological information that has historically been confined to a small community of experts.

725. The parents of Hannah Bruesewitz sued Wyeth Laboratories, now owned by Pfizer, Inc., claiming that their daughter suffers a "residual seizure disorder" as a result of receiving the diphtheria, pertussis, and tetanus (DPT) vaccine, and alleging that the company failed to adequately warn them of risks associated with the vaccine.

726. The only concern with Imetrix is that any patient with any cardiovascular issues is warned from taking Imetrix due to risks associated with the dialation of the blood vessels in a patient taking Imetrix.

727. Furthermore, the veteran population may have a higher prevalence of factors associated with the development and progression of OSA, such as excess body weight, smoking, alcohol consumption, and nasal congestion (1).

728. This shortage resulted from termination of operations by two major producers of fluoride products, realignment of other producers, and factors associated with the globalization of the chemical industry.

729. They also looked at factors associated with the preservation of B cell function.

730. They speculate that "the higher rate of death in the intensive-therapy group may be related to factors associated with the various strategies.

731. Johnston County Osteoarthritis Project This ongoing community-based cohort study of rural white and black persons to determine the prevalence, incidence, and factors associated with the occurrence or progression of hip and knee osteoarthritis includes a genomics component for examining genes (HH, COMP, COL2A, others to be determined) that may be linked to osteoarthritis and related conditions.

732. Veterans may be at an elevated risk for OSA because of increased prevalence of factors associated with the development and progression of OSA.

733. No patients above the age of 65 years were enrolled in double-blind prospective clinical trials of mania associated with bipolar illness using DEPAKOTE (divalproex sodium delayed-release tablets).

734. The FDA had previously approved Zyprexa to treat acute episodes of mania associated with bipolar disorder.

735. No patients above the age of 65 years were enrolled in double-blind prospective clinical trials of mania associated with bipolar illness.

736. There are a range of symptoms associated with fibromyalgia, including chronic widespread pain and tenderness.

737. Because of the diversity of symptoms associated with fibromyalgia, your trainer should also be able to properly progress and regress exercises based on your needs.

738. This can be tricky, however, because symptoms associated with fibromyalgia can be caused by other conditions.

739. When properly used, these cleaning, disinfection, and sterilization processes can reduce the risk for infection associated with use of invasive and noninvasive medical and surgical devices.

740. The choice of disinfectant, concentration, and exposure time is based on the risk for infection associated with use of the equipment and other factors discussed in this guideline.

741. Fatigue associated with depression may occur with excessive crying, insomnia and apathy.... Read more about fatigue symptoms Fatigue can be caused by a wide variety of diseases, disorders or conditions, such as anemia, low blood pressure (hypotension), chronic fatigue syndrome, and Addison's disease.

742. Fatigue associated with depression may occur with excessive crying, insomnia and apathy.

743. Although there are no case reports to indicate cross sensitivity with other drugs that produce this syndrome, the experience amongst drugs associated with multi-organ hypersensitivity would indicate this to be a possibility (see WARNINGS, Patients with a Past History of Hypersensitivity Reaction to Carbamazepine subsection).

744. Although the existence of cross sensitivity with other drugs that produce this syndrome is unclear, the experience amongst drugs associated with multi-organ hypersensitivity would indicate this to be a possibility.

745. Although antidepressant drugs were developed to treat depression, it has been discovered that they are also effective in combating chronic headaches, cancer pain, and pain associated with nerve damage.

746. Although antidepressant drugs were developed to treat depression, they are also effective in combating chronic headaches, cancer pain, and pain associated with nerve damage.

747. Sotalol hydrochloride (AF) should be used only with extreme caution in patients with sick sinus syndrome associated with symptomatic arrhythmias, because it may cause sinus bradycardia, sinus pauses or sinus arrest.

748. Sotalol hydrochloride tablets (AF) should be used only with extreme caution in patients with sick sinus syndrome associated with symptomatic arrhythmias, because it may cause sinus bradycardia, sinus pauses or sinus arrest.

749. The symptoms of childhood dermatomyositis are similar to those associated with the adult form of the disorder.

750. The NIOSH Alaska field station is charged with collecting information on all occupational fatalities, focusing on those associated with the fishing, logging, and air transport industries.

751. However, there are only a few kinds of pain associated with MS; others should be considered suspicious.

752. Another type of pain associated with MS is musculoskeletal pain which occurs in the muscles, tendons, and ligaments around joints.

753. Experimental inputs will be used to drive the model such that the forces and stresses of the internal structures (e.g., cartilage, ligaments, meniscus) may be evaluated for postures associated with low-seam mining.

754. These kneel-assist devices will more effectively reduce the forces, moments, and stresses applied to the knee while in postures associated with low-seam mining compared to currently available kneepads.

755. We will be able to use this model to evaluate the forces in the ligaments while in postures associated with low-seam mining.

756. Radiofrequency ablation, also known as RFA, can help patients with chronic lower back and neck pain, and pain associated with the degeneration of joints from arthritis.

757. The specific surgery that is right for you depends on the extent and pattern of cartilage damage and level of pain associated with the ankle.

758. Surprisingly, there is not a lot of pain associated with the surgery.

759. Although I found these to be quite painful they worked pretty well and I was more then willing to take 5 minutes of pain verse hours and days of pain associated with the migraines.

760. Systemic effects such as nausea, headache, and increased heart rate appeared to be associated with DMEA concentrations higher than those associated with blurred vision only [Stephenson and Albrecht 1986].

761. Acute physical distress and halo vision were experienced at concentrations higher than those associated with blurred vision only.

762. Complications associated with urine color changes vary depending on the underlying disease, disorder or condition.

763. Complications associated with urine odor vary depending on the underlying disease, disorder or condition.

764. The clinical characteristics of the seizures reported in patients with serum theophylline concentrations ¡ 20 mcg/mL have generally been milder than seizures associated with excessive serum theophylline concentrations resulting from an overdose (i.e., they have generally been transient, often stopped without anticonvulsant therapy, and did not result in neurological residua).

765. The clinical characteristics of the seizures reported in patients with serum theophylline concentrations ¡20 mcg/mL have generally been milder than seizures associated with excessive serum theophylline concentrations resulting from an overdose (i.e., they have generally been transient, often stopped without anticonvulsant therapy, and did not result in neurological residua).

766. The clinical characteristics of the seizures reported in patients with serum theophylline concentrations ¡20 mcg/mL have generally been milder than seizures associated with excessive serum theophylline concentrations resulting from an overdose (i.e. they have generally been transient, often stopped without anticonvulsant therapy, and did not result in neurological residua).

767. National Children's Center for Rural and Agricultural Health and Safety This Center funded by NIOSH and the Federal Maternal and Child Health Bureau strives to enhance the health and safety of all children exposed to hazards associated with agricultural work and rural environments.

768. National Children's Center for Rural and Agricultural Health and Safety The National Children's Center for Rural and Agricultural Health and Safety strives to enhance the health and safety of all children exposed to hazards associated with agricultural work and rural environments.

769. NCCRAHS strives to enhance the health and safety of all children exposed to hazards associated with agricultural work and rural environments.

770. The Center conducts research, education, intervention, prevention, translation and outreach activities to enhance the health and safety of children exposed to hazards associated with agricultural work and rural environments.

771. The overall mission of the National Children's Center for Rural and Agricultural Health and Safety is to enhance the health and safety of all children exposed to hazards associated with agricultural work and rural environments.

772. It affects both feet, my ankle, left hip, leg, spine, wrist and jaw bone.

773. It affects both women and men, and especially MSers.

774. It affects both sexes equally.

775. It affects both sexes approximately equally.

776. It affects both sexes but is more common in women and may begin at any age, even in young children.

777. (Created: 4/18/2008 by NCHHSTP, date released: 5/1/2008, running time: 6:22) Stop the Belly-Aching (Rotavirus) Rotavirus, an illness characterized by diarrhea, vomiting, and dehydration, affects nearly every child aged less than five years.

778. Stop the Belly-Aching (A Cup of Health with CDC) April 2008 Listen to this podcast (3:05) Rotavirus, an illness characterized by diarrhea, vomiting, and dehydration, affects nearly every child younger than 5 years of age.

779. Inner ear decompression sickness, a condition that affects scuba divers, usually occurs within 30-60 minutes after a dive.

780. Inner ear decompression sickness (IEDCS) is a serious condition that affects scuba divers.

781. While it's most common in those over age of 45, it affects many younger adults as well, often triggered by a work-, accident-, or sports-related injury.

782. Although ADHD is often considered a child's disorder it affects many adults as well.

783. First, the disease burden is high (i.e., it affects many people, has increased recently, and will likely increase in the future).

784. HPV affects both men and women and can cause warts, or papillomas, on ... Read more about human papillomavirus introduction Human papillomavirus (HPV) is a contagious virus transmitted through direct contact, including sexual contact.

785. HPV affects both men and women and can cause warts, or papillomas, on the genitals and around the anus, as well as at other sites.

786. In addition, he has observed a lot about how food affects your health.

787. Bottom line: 1. Eat when hungry 2. Stop when full 3. Be aware when you turn to food for emotional reasons 4. Discover how food affects your hunger and fullness, ie higher fiber foods keep you satisfied longer 5. Once you have that down, fine tune your nutrition to optimize health.

788. It affects an estimated 1.5 to 3.5 million school-age children in the U.S. Everyone, especially younger children, may have symptoms of ADHD from time to time.

789. It affects an extremely broad range of birds, it affects an extremely broad range of mammals, and it affects an extremely broad range of mosquitoes.

790. Are you testing to see what affects you?

791. Finding what affects you most can help you be prepared to address the situations when they arise.

792. An emetic is "biologically appropriate" for the patient in that it affects him or her in the same organ systems that excessive alcohol use does.

793. So I never forget that IBD doesn't just affect me, but it affects him as well.

794. I need to keep in mind, however, that each person is different and moreover, MS affects each person differently.

795. Although MS affects each person differently, the disease generally occurs in one of four patterns or clinical courses, which are sometimes referred to as chronic progressive MS.

796. Testosterone Testosterone is a hormone which affects sexual features and development.

797. Testosterone is a hormone which affects sexual features and development.

798. Increasingly however there is evidence to show differences in the causes and expression of depression and the way it affects lives.

799. Together, the 10 banners and their firstperson accounts told a compelling story about CFS and how it affects lives.

800. Dysautonomia- it affects the spine.

801. The distinguishing characteristic of mucosal melanoma is that it affects the body's mucous membranes.

802. Prolactinoma can affect both men and women, but it affects the sexes somewhat differently.

803. In general, researchers believe that congenital muscular dystrophy has such an early onset because it affects the muscle proteins that are required for the development of muscles that are especially important during infancy.

804. How it affects the body, some of the symptoms, medications that are used to treat erectile dysfunction and treatments available.

805. But any damage is of concern because we don't know enough about how it affects the whole body.

806. Ringworm or tinea is considered a kind of fungal infection and it affects the hair, skin, and nails.

807. We really don't know what causes depression or how it affects the brain.

808. Diphtheria is especially dangerous when it affects the throat, where it can produce a thick gray membrane that may grow large enough to obstruct breathing.

809. The type of pain a person feels depends on the type of cancer the person has and how it affects the body.

810. This is an excellent article underlining the stress and how it affects the health.

811. I have read alot of articles about myasthenia gravis and they all state that it affects the nerves and muscles and that the rest is the best thing for you.

812. Neurosyphilis is different from syphilis because it affects the nervous system, while syphilis is a sexually transmitted disease with different signs and symptoms.

813. In most cases, bladder outlet obstruction is important only in as much as it affects the person's quality of life.

814. Anyone with a long-term chronic disease like asthma can always learn more about what asthma is, how it's caused, how it's diagnosed, how it affects the body and how it's treatable at any age.

815. I mean we're talking about economic, you know, the socioeconomics of African Americans, been talking about poverty and how it affects the whole thing.

816. Studies have... Read more A Migraine attack is not just a bad headache; rather, it affects the entire body.

817. In about 40 percent of people who have Raynaud's, it affects the toes.

818. Besides the great psychological repercussions, for the diabetic patient and his family; because it affects the social, physical and emotional part of the family environment.

819. Chondrocalcinosis is also called pseudogout or pseudo-osteoarthritis, the latter particularly when it affects the knees.

820. Doctors use a four-level grading system to categorize astrocytomas, and the system helps describe how various types of astrocytomas affect the body: Whether a brain tumor grows slowly or quickly and whether its borders are well-defined or not influence how it is treated and also make a difference in how it affects the body.

821. When it is severe, it affects the mother's circulatory system, kidneys, brain and other vital organs.

822. If your child senses your stress, it affects the outcome!

823. And since it affects the brain, it naturally affects much of what we do, say, perceive and think about ourselves and others.

824. Tinea is considered a kind of fungal infection and it affects the hair, skin, and nails.

825. If it affects the brain and spinal cord, people may have changes in mood and behavior and may even become paralyzed, so it's important to understand these effects and to support them physically and emotionally in whatever ways you can.

826. You can read what bipolar is, how it affects the person who has it and how it affects those around them.

827. So, I hope you knowing about me will help don't give up there are people that do like to go get drugs from doctors, and unfortunately it affects the true patients that have extreme pain such as us.

828. If someone you know suffers from it do know that although it affects the skin it can also affect many facades of a person.

829. An important element of this task will be to explore the community's legal, policy, and social environment and how it affects the services and programs available to IDUs.

830. When it affects the coronaries mainly as blockages in the arteries leading to angina or heart failure or event heart attack it is known as coronary heart disease.

831. As with disseminated chickenpox, disseminated herpes zoster, which spreads to other organs, can be serious to life-threatening, particularly if it affects the lungs.

832. The only type of tuberculosis that is contagious is the active variety, when it affects the lungs.

833. ROM may be the most important because it affects the patient's ability to use the knee.

834. Most commonly, though, it affects the small intestine or the colon or both.

835. When cortisol becomes elevated and remains so for awhile, it affects the cells that comprise your immune system.

836. These conditions include: Anterior uveitis is often referred to as iritis because it affects the iris.

837. Sometimes, it affects the way the thyroid functions.

838. IP is often referred to as a neurocutaneous condition because it affects the nervous system and the skin.

839. In humans, dichloroacetylene exposure causes headache, loss of appetite, extreme nausea, and vomiting; it affects the trigeminal nerve and facial muscles and exacerbates facial herpes.

840. The extent of amniotic fluid deficiency in oligohydramnios and the trimester it occurs in can both influence how seriously it affects the mother and her fetus.

841. If part of the stomach pushes through the opening at the bottom of the diaphragm (hiatus) or a weak area around the hiatus, it affects the digestive system.

842. Regardless of its cause and how it affects the body, men with infertility may be affected emotionally and psychologically.

843. You can also read more about anxiety and how it affects the body here in the overview.

844. If it affects the heart, it may lead to heart failure and other heart disease.

845. In most cases, amyloidosis is a systemic disease, meaning that it affects the entire body.

846. The main difference is, it affects the brain itself and cognitive function.

847. Because it affects the nervous system, TSD is classified as a neurological disease.

848. Because it affects the vulva, however, they do not think of it as a dermatologic issue.

849. If it affects the nervous system, syphilis can cause mood and behavior changes, stroke, and even paralysis.

850. Now I'm very reluctant to take anything new especially if it affects the neurotransmitters, and almost all of the new ones do.

851. As dysprosody is such a rare condition, researchers are still trying to pinpoint exactly how it affects the body.

852. Posterior uveitis may also be referred to as choroiditis because it affects the choroid.

853. Cardiovascular disease (CVD) is the term used to describe atherosclerosis, or hardening of the arteries, as it affects the heart.

854. Not only does it affect the person with the diagnosis, it affects the spouse, adult children and even the grandchildren.

855. A Migraine attack is not just a bad headache; rather, it affects the entire body.

856. Like most chronic conditions, it affects the whole family.

857. That means any small change in my dosage could make a big difference in how it affects the blood.

858. It should be accepted for what it is and how it affects the person.

859. it's not really that prednisone is working on the brain per se, but it's just stopping the immune system from reacting so strongly that it affects the brain's function.

860. Hypogonadism can affect both men and women, but it affects the sexes somewhat differently.

861. I hate the label and it affects the pills doctors will prescribe me or won't prescribe me (it also makes me unqualified for all of the new innovative treatments like neuromodulation that are 'reserved' for unipolar depressives).

862. For instance, Marfan syndrome can be life threatening if it affects the aorta, the main artery that carries blood from the heart to the rest of the body.

863. The distinction between nerve and vascular problems is an important distinction to make because it affects the treatment.

864. Here is a great overview of what being diagnosed with Anxiety actually means, how it affects the mind and body.

865. This is more of an irritation when it is a superficial fungal infection, but very serious when it affects the deep tissues and organs.

866. Neurological issues: When skeletal support is compromised, it affects the whole body.

867. While this is certainly a topic for debate as to who obesity affects most, my opinion on the matter is that it affects the patient the most.

868. I recollect your clarifications on the scary topic of hyper parathyroidism and hypercalcemia and the way it affects the body, and how you advised my sister in India.

869. When it affects the small intestine, lymphangiectasia may be caused by another illness.

870. This work will provide the ground work for future research examining how proprioception is altered by occupational vibration exposure and how it affects the overall spinal stabilization.

871. If you are pregnant or nursing, do not use St. Johns Wort unless you consult your health care professional, since it is unknown if it affects the unborn or babies.

872. When an organ, usually a portion of the intestine, slips through the abdominal wall and into other areas of the body, it affects the organs and structures it encounters.

873. It affects us all.

874. It affects us emotionally, physically, and behaviorally.

875. While drinking water won't actually moisturize your skin, it will help your internal organs function properly, which affects your skin's health and beauty.

876. Love to all of you, wishing all the best, Linda I have carcinoid cancer, which affects your mood and often I become depressed.

877. Renin works with aldosterone (a hormone made by the adrenal glands) and several other substances to help balance sodium and potassium levels in the blood and fluid levels in the body, which affects your blood pressure.

878. Unfortunately it also takes away depth perception which affects your driving, too, so be careful.

879. It is also possible to experience internal gangrene, which affects your inner tissues or organs.

880. Chemo can also cause a temporary pause in menstrual cycles which would cause a loss in estrogen which affects your bones.

881. Do you have a chronic skin condition which affects your mental health?

882. If you know of a potential trigger that affects you, you can prevent symptoms by avoiding that trigger.

883. This is strong stuff that affects you mentally as well as physically.

884. Genetic counseling can help you understand the inheritance pattern of the type of Ehlers-Danlos syndrome that affects you and the risks it poses for your children.

885. Just as I am grateful for your work that affects you & so many more of us.

886. I had an mri which was normal, (they checked me for MS) This happens at least once a year and now i am getting fearful because of how long this affects me.

887. It's not like this affects me every moment, minute by minute, of every day, thank God!!!

888. Arthritis affects 50 million Americans and is expected to increase significantly as the population ages.

889. Big numbers: Arthritis affects 50 million adults, limiting the activities of nearly 21 million, and is the most common cause of disability in the United States.

890. The most common of the inherited peripheral neuropathies in the United States is Charcot-Marie-Tooth disease, which affects approximately 125,000 persons.

891. No one knows the specific causes of type 1 diabetes, which affects approximately 5 - 10 percent of all those diagnosed with diabetes.

892. TB can affect the bones, joints, lymph nodes, kidneys, liver, and brain, but the most common form is pulmonary TB, which affects the lungs.

893. Together, the studies included six patients with genetic mutations that cause Leber's congenital amaurosis, which affects the retina.

894. Multiple Sclerosis is a neurological disease which affects the nervous system.

895. Infection by Haemophilus influenzae type b (Hib) bacteria can cause life-threatening illnesses, such as meningitis, which affects the brain; epiglottitis, which affects the throat and can cause death by suffocation; pericarditis, which affects the heart; pneumonia, which affects the lungs; and septic arthritis, which affects the bones and joints.

896. These are known as poststreptococcal glomerulonephritis, which affects the kidney, and rheumatic fever, which affects the heart.

897. These signs and symptoms are characteristic of Hartnup disease, which affects the skin and the brain.

898. Examples of organ-specific autoimmune disorders are insulin-dependent diabetes (Type I) which affects the pancreas, Hashimoto's thyroiditis and Graves' disease which affects the thyroid gland, pernicious anemia which affects the stomach, Addison's disease which affects the adrenal glands, and chronic active hepatitis which affects the liver.

899. Cervicofacial actinomycosis, which affects the jaw, face, and neck, is the most common type of actinomycosis infection, accounting for 50-70% of all cases.

900. Fungus usually grows when your nails are continually exposed to moist environments, like sweaty shoes or shower floors (by the way, it's not the same as athlete's foot, which affects the skin).

901. Pneumococcal infection can cause serious problems, such as pneumonia, which affects the lungs; meningitis, which affects the brain; bacteremia, which is a severe infection in the blood; and possibly death.

902. Bifenthrin is an insecticide and acaricide which affects the nervous system and causes paralysis in insects, quickly killing all tick species including dog and deer ticks.

903. These groups cause approximately 50% of meningococcal meningitis cases in the U.S. The vaccine will not protect against infection caused by other meningococcal bacteria groups, such as Group B. Meningococcal infection can cause life-threatening illnesses, such as meningococcal meningitis, which affects the brain, and meningococcemia, which affects the blood.

904. I'm 48 and having chemotherapy treatment which affects the ovaries, puts you into almost sudden menopause which is permanent for 90% of women in my age group.

905. At dosages greater than required for beta blockade, propranolol also exerts a quinidine-like or anesthetic-like membrane action, which affects the cardiac action potential.

906. Meningococcal infection can cause life-threatening illnesses, such as meningococcal meningitis, which affects the brain, and meningococcemia, which affects the blood.

907. Phyllis Kanki, an AIDS specialist and professor of immunology and infectious diseases at the Harvard School of Public Health, suggested that people who get infected with HIV-2, which affects the body more slowly, may develop better defenses against the virus.

908. Infection by Haemophilus influenzae type b (Hib) bacteria can cause life-threatening illnesses, such as meningitis, which affects the brain; epiglottitis, which can cause death by suffocation; pericarditis, which affects the heart; pneumonia, which affects the lungs; and septic arthritis, which affects the bones and joints.

909. The hormones – progesterone and estrogen – also cause your muscles to relax, which affects the digestive tract muscles.

910. Acne is a common skin disorder that occurs in two forms: superficial (acne vulgaris), which affects the hair follicles and oil-secreting glands of the skin and manifests as blackheads, whiteheads, and inflammation, and cystic (acne conglobata), a more severe form, with deep cyst formation and subsequent scarring.

911. The more advanced the cancer upon detection, the greater likelihood that chemotherapy, which affects the whole body, will be used to treat it.

912. The most common type is peripheral neuropathy, which affects the arms and legs.

913. During this stage of life, a womans body makes less of the sex hormone estrogen, which affects the hypothalamus (this gland regulates your body temperature).

914. These children experience a developmental defect, typically during the first trimester, which affects the spinal cord, the fluid-filled sac that surrounds the spinal cord, and sometimes the surrounding nerves.

915. About 3 percent to 5 percent of women who get endometrial cancer, which affects the womb lining, are under the age of 40 and will lose their fertility if they undergo a hysterectomy.

916. Another is anencephaly, which affects the brain and results in miscarriage, stillbirth, or babies who live only a few days.

917. The bacteria then multiply and produce a toxin, which affects the nervous system.

918. Varenicline, which affects the brain-nicotine relationship and decreases both withdrawal symptoms and cravings, must be used alone.

919. Other diseases causing back pain include arthritis, which erodes the joints, myopathies and inflammatory conditions, which involve the muscles, and neuropathy, which affects the nerves.

920. Maybe you do have a demyelinating disease such as multiple sclerosis (MS) which affects the central nervous system or chronic inflammatory demyelinating polyneuropathy (CIDP) which affects the peripheral nervous system.

921. And in recent clinical trials, the antidepressant bupropion (Wellbutrin), which affects the brain chemical dopamine, showed benefits for adults with ADHD.

922. These groups cause nearly all of the meningococcal meningitis cases in the U.S. The vaccine will not protect against infection caused by other meningococcal bacteria groups, such as Group B. Meningococcal infection can cause life-threatening illnesses, such as meningococcal meningitis, which affects the brain, and meningococcemia, which affects the blood.

923. Dandruff is a common skin condition which affects the scalp and causes flaking skin.

924. In dosages greater than required for beta blockade, propranolol also exerts a quinidine-like or anesthetic-like membrane action, which affects the cardiac action potential.

925. Essentially, the autonomic nervous system no longer functions correctly which affects the organs.

926. Microvascular angina may be an indicator of coronary artery disease, which affects the large arteries.

927. With anemia, the body has low levels of iron, which affects the blood's ability to carry oxygen throughout your body.

928. At doses greater than required for beta blockade, propranolol also exerts a quinidine-like or anesthetic-like membrane action, which affects the cardiac action potential.

929. Only about 2,000 new cases are diagnosed each year of this type of cancer, which affects the mesothelium – a membrane that covers and protects most of the body's internal organs.

930. Niacin deficiencies cause a wasting disease known as pellagra, which affects the skin, mucous membranes, gastrointestinal tract as well as the brain, spinal cord and peripheral nerves.

931. Cervical dystonia, which affects the head and neck, is the most common adult form of dystonia, followed by blepharospasm (eyelids), spasmodic dysphonia (larynx), and limb dystonias (hands).

932. Anything which affects the nerves could cause numbness: disc compression in the spine, diabetes, multiple sclerosis, etc., etc.

933. Common eye infections include conjunctivitis, often called pink eye, which affects the membrane that lines the inside of your eyelids and covers the whites of the eyes, and blepharitis, which affects the eyelid margin.

934. Viruses attack the immune system, which affects the hormonal system.

935. In dosages greater than required for beta-blockade, propranolol also exerts a quinidine-like or anesthetic-like membrane action, which affects the cardiac action potential.

936. At dosages greater than required for beta-blockade, propranolol also exerts a quinidine-like or anesthetic-like membrane action, which affects the cardiac action potential.

937. Stressful situations over which a person has no control appear to activate a brain enzyme called protein kinase C (PKC), which affects the prefrontal cortex.

938. Rheumatoid arthritis is caused by a system-wide disease process, which affects the entire body.

939. Another cause of claw hand is the bacterial disease leprosy, which affects the skin.

940. Thiamine deficiency, or beriberi, manifests itself as both wet beriberi, which affects the cardiovascular system, and dry beriberi, which causes neurological dysfunction.

941. Thoracic actinomycosis, which affects the chest cavity, makes up about 15-20% of actinomycosis cases.

942. Currently, there are no tests that can reliably detect either ovarian or endometrial cancer, which affects the uterine lining.

943. Each laser produces one specific color, which affects the body differently.

944. Means the joint fluid doesn't have enough hyaluronan, which affects the joint's ability to absorb shock.

945. When lymphomas begin, malignant lymphocytes multiply uncontrollably and do not perform their normal functions, which affects the body's ability to fight infections.

946. Based on the recommended calorie intake for men and women at certain age levels, the RDAs for thiamine are: A disease called beriberi, which affects the nerves and heart, is caused by a lack of thiamine in the diet.

947. Dental fluorosis is a concern that affects primarily children.

948. More It has been known for several years by the analysis of many large families, that ND is an inherited condition that affects primarily males.

949. It has been known for several years by the analysis of many large families, that ND is an inherited condition that affects primarily males.

950. I do understand that an illness that affects mood may be hard to imagine for some.

951. Depression is an illness that affects mood, body, behavior and mind.

952. It quickly circulates to the brain where it indirectly increases the supply of dopamine, a chemical in the brain that affects mood.

953. But heat is not the only thing which affects my MS negatively.

954. I become extraordinarily fatigued which affects my physical and cognitive abitilies.

955. In children, it affects boys somewhat more often than girls.

956. It's more common in women and girls, but it affects boys and men, too.

957. It affects everything, it seems.

958. It affects everything.

959. Baby bottle tooth decay (BBTD) is a type of early childhood caries (ECC) that affects primary teeth.

960. Pre-pubertal periodontitis, which begins before puberty and that affects primary teeth, is extremely rare.

961. It is generally a disease that affects young people, but doctors don't understand yet what causes it.

962. Among these are osteosarcomas (osteogenic sarcomas) and Ewing's sarcoma, a particularly aggressive tumor that affects young adults.

963. It affects most body tissues, particularly bones, teeth, and blood vessels.

964. It affects most women two to three days before their period begins and can last throughout the period.

965. It affects most people on the left side that you mentioned, and untreated can cause fever and other problems.

966. Remember that MS affects everyone differently so what one person with MS experiences may not be what you will experience.

967. MS affects everyone different .

968. Depression affects nearly one in six people at some point in their lives, so folk remedies and half-truths about this common illness abound.

969. Depression affects nearly 18.8 million American adults each year, including persons of all income levels, educational backgrounds, and professions.

970. Not surprisingly, this affects our energy, our cognitive function, motivation, moods, personalities, and feeds into our illness.

971. I know the break down of it and why double bonds can not twist because of the orbitals, however, I am confused as to how and why this affects our diet?

972. For example, the fetus may have abnormal chromosomes, or a genetic defect, that affects health and development.

973. Some involve stations displaying an

974. or activity that affects health.

975. This bacterium makes a toxin that can cause infant botulism a form of food poisoning that affects a baby's nervous system and can result in death.

976. Hypertrophic cardiomyopathy (HCM or HCOM) is a disease that affects a person's heart muscle, making the muscle more thick and rigid than it should be.

977. With a little tweaking, the authors said the task used to measure infants" responses might be adapted to help with the early detection of autism, a condition that affects a child"s ability to communicate and socialize.

978. Cancer of the uterus is the most frequent and most curable type of cancer that affects a woman's reproductive system.

979. Shyness is a personality trait that affects a child's temperament.

980. What it means: It's not entirely clear what it is about iron-deficiency anemia that affects a chid's respiratory development, says Elizabeth Triche, PhD, assistant professor of epidemiology at Brown University and the study's lead author.

981. Dementia is a deterioration in brain function that affects a person's ability to think, reason, and feel emotions.

982. Reality: SZ is a no-fault biological brain disorder that affects a person's moods, thoughts and behavior, sometimes causing us to act in bizarre ways.

983. So important that dignity and respect for human life, regardless of the disease that affects a person, remain at the forefront.

984. "It isn't just sugar," that affects a diabetic, said Dr. Harvey Katzeff, chief of the endocrinology division at North Shore Long Island Medical Center in New York.

985. The second most common type is Frontotemporal Disease, a dementia that affects a younger population.

986. While seeking or negotiating for a non-government job, you cannot work on a matter at the Department that affects a prospective employer.

987. It provides mental, physical, emotional, and other benefits for anyone healing from lung... Read more Any disorder that affects a person's lungs is referred to as lung disease.

988. Gentrification is a housing, economic, and health issue that affects a communitys history and culture and reduces social capital.

989. Neurological disorders: People with LNS tend to have poor muscle tone, which may cause coordination problems that resemble cerebral palsy, a disorder that affects muscle tone, movement, and motor skills.

990. Acid maltase deficiency (AMD), also known as Pompe disease, is a genetically inherited disease that affects muscle function.

991. It was so easy to blame the meds, but then I got serious about my physical health and how it affects my mental health, and I have lost 90 lbs! Yes, I'm still on the same meds that I blamed for the weight gain, and the pot, too.

992. In my case, it affects my arms, so I'm very limited in what I can do with my upper body.

993. In the 3 years, the only drug I've been able to stick with is Arava, but I can only take it 4 times a week as it affects my liver too.

994. I eat in moderation and watch which carbs I put in my mouth and how it affects my blood sugar.

995. I'm just confused as to why it affects my upper back and am not really convinced that's what it is.

996. So what the heck does this have to do with MS and how it affects my life??

997. i cant lay on my side because of my shoulders, and i cant lay on my back because it affects my back and hurts my hips, when i lay on my frount sometimes my back locks or i cant get up.

998. I am self-conscious and it affects my confidence and sense of health.

999. And then it affects my comprehension.

1000. when it affects my hands, elbows and shoulders, I can't pick up my baby.

1001. I thought I was having troubles with bad memory and how it affects my family but all I'm thinking about right now is your struggles.

1002. One thing for sure; I feel guilty on a regular basis because my illness causes me to function at a lower level with daily activities than I would like, it affects my family and I HATE IT.

1003. He knows I'm bipolar, but I dont think he has a clue as to how much it affects my life.

1004. Who cares if it affects my career or other parts of my life?' Let me start by confirming that the Cosmo survey did not look at how relative weight affects or influences married folk.

1005. The MS hit first and harder on my right leg; it affects my left leg far less.

1006. What i mean is I feel like it affects my children now that they are teenagers.

1007. I think that it must be part of the disease for me...in how it affects my mental health and body chemistry.

1008. I have been in a long relationship now and my significant other knows all about me, the diabetes and how it affects my life.

1009. The whole "feet" thing is scary, since it affects my mobility.

1010. Stress affects everyone differently and that includes how it may impact your eating habits and appetite.

1011. Stress affects everyone but people with TBI are particularly vulnerable to this problem.

1012. Osteoarthritis, the wear-and-tear form of arthritis, affects one in two Americans during the course of their lifetime.

1013. When arthritis affects one part of the knee, rather than the whole knee, this is unicompartmental arthritis.

1014. Stress affects the brain too, causing changes in its structure and connectivity.

1015. Stress affects the mind, body, and behavior in many ways, and everyone experiences stress differently.

1016. According to the World Health Organization, depression affects 121 million people worldwide.

1017. The World Health Organization estimates that depression affects 121 million people worldwide.

1018. It affects my work and my social life.

1019. It affects my head.

1020. It affects my life daily, even when I don't have an attack.

1021. It affects my every day life.

1022. It affects my work, my school, my life.

1023. It affects my relationships, my work my whole life, Its hard to be pleasant when my head is split by a knife.

1024. It's a wonder to see the interaction and how it affects both sides.

1025. Chronic fatigue syndrome can have serious affects on someone's daily functioning because it affects both body and mind.

1026. In contrast, damage to the optic nerve chiasm or the pathway beyond it affects both eyes.

1027. If it affects both breasts, it's almost certainly an allergy, or a hormonal issue.

1028. Cymbalta (duloxetine) is a new antidepressant that is classified as a selective serotonin and norepinephrine reuptake inhibitor (SSNRI) because it affects both neurotransmitters.

1029. This disease usually affects only one knee, though occasionally it affects both knees.

1030. This isn't a breast cancer symptom, seeing as it affects both breasts and has been going on for so long.

1031. Although it affects both healthy cells as well as cancer cells, the rapidly dividing cancer cells are more susceptible to the drug's effect.

1032. This little known disease is the fourth deadliest cancer and it affects both mean and women equally.

1033. "Depression occurs disproportionately among new parents, and this study also hammers home the point that it affects both mothers and fathers," says James F. Paulson, PhD, associate professor of pediatrics at the Eastern Virginia Medical School, Norfolk.

1034. Patients should be advised that they should neither drive a car nor operate other complex machinery until they have gained sufficient experience on TASMAR to gauge whether or not it affects their mental and/or motor performance adversely.

1035. Some people are seeking information about a disorder or its treatment, many people are giving or receiving emotional support and encouragement and some are venting frustrations about their condition or how it affects their life.

1036. The Brief Pain Inventory looks at the patients' pain history, intensity, and location, as well as how much it affects their daily life.

1037. Since somnolence is a frequent adverse event with potentially serious consequences, patients should neither drive a car nor engage in other potentially dangerous activities until they have gained sufficient experience with APOKYN to gauge whether or not it affects their mental and/or motor performance adversely.

1038. For Betty — remember how Alzheimer's works and how it affects their brains.

1039. If your loved one wants to talk about the effects of prolactinoma - especially if it affects their fertility - be open to it, respect their feelings, and support them emotionally however you can.

1040. I wish people could see how much this affects the adults and childred who have this and how much it affects their family and friends maybe if they walked in our shoes for a few miles they might learn to understand that this is not just a lack of discipline .

1041. Accordingly, they should be advised neither to drive a car nor to operate other complex machinery until they have gained sufficient experience on gabapentin to gauge whether or not it affects their mental and/or motor performance adversely.

1042. Denise I agree that more people with MS need to be stepping forward and sharing what it is like and how it affects their daily lives.

1043. I recommend FM patients to buy a book on leaky gut to find out themselves how it affects their health.

1044. Researchers from Penn State say while they know testosterone makes males more susceptible to disease, they also wanted to understand if it affects their behavior and how that increases their ability to transmit disease.

1045. But now researchers are concerned that men who take daily doses of painkillers such as aspirin and ibuprofen may be lowering their PSA levels to the point that it affects their doctors' ability...

1046. Patients should be warned about the potential for dizziness, vertigo, or somnolence and advised not to drive or operate machinery until they have gained sufficient experience on RILUTEK to gauge whether or not it affects their mental and/or motor performance adversely.

1047. Accordingly, they should be advised neither to drive a car nor to operate other complex machinery until they have gained sufficient experience on Neurontin to gauge whether or not it affects their mental and/or motor performance adversely.

1048. Since somnolence is a frequent adverse event with potentially serious consequences, patients should neither drive a car nor engage in other potentially dangerous activities until they have gained sufficient experience with REQUIP to gauge whether or not it affects their mental and/or motor performance adversely.

1049. They lose a lot of days from work, it affects their social life, and they are always looking for bathrooms, he says.

1050. When it does, it affects their shape and how they work.

1051. Syphilis is a type of sexually transmitted disease (STD) caused by an infection of the bacterium Treponema pallidum.

1052. Symptoms begin to appear about six weeks after the chancre has resolved and include: Syphilis is a type of sexually transmitted disease (STD) caused by an infection of the bacterium Treponema pallidum.

1053. Gamboge is a type of laxative called a stimulant laxative.

1054. Gamboge is a type of laxative that might also decrease potassium in the body.

1055. Intussusception is a type of bowel blockage caused when the bowel folds into itself like a telescope.

1056. Intussusception is a type of bowel blockage that is treated in a hospital.

1057. This is a type of breast cancer which is estimated to be 10%-20% of all breast cancers.

1058. This is a type of breast infection that fits your description.

1059. Chronic bronchitis, like emphysema, is a type of chronic obstructive pulmonary disease (COPD).... Read more about bronchitis causes Treatment for bronchitis begins with seeking medical care from your health care provider.

1060. Chronic bronchitis, like emphysema, is a type of chronic obstructive pulmonary disease (COPD).

1061. Shingles is caused by the varicella-zoster virus, which is a type of herpes virus.

1062. Mononucleosis is a viral infection usually caused by the Epstein-Barr virus (EBV), which is a type of herpes virus and one of the most common human viruses.

1063. Neuroblastoma is a type of cancer that usually originates either in the tissues of the adrenal gland or in the ganglia of the abdomen or in the ganglia of the nervous system.

1064. Neuroblastoma is a type of cancer found mostly in young children.

1065. Dermabrasion is a type of surgery that can remove surface scars and reduce the depth of deep scars.

1066. Dermabrasion is a type of surgery done with a rotating brush used to treat various skin conditions.

1067. Pachyonychia congenita with natal teeth: This is a type of pachyonychia congenita (elephant nails from birth) in which teeth are evident at birth.

1068. Pachyonychia congenita of the Jadassohn-Lewandowski type: This is a type of pachyonychia congenita (elephant nails from birth).

1069. Mold is a type of fungus.

1070. Mold is a type of fungus, a diverse kingdom that also includes mushrooms and yeasts.

1071. Secondly, Alzheimer's is a type of dementia (the most common, it is thought), but there are many types of dementia.

1072. Firstly-Alzheimer's is a type of dementia.

1073. Some people still ask what the difference between Alzheimer's and dementia is (Alzheimer's is a type of dementia).

1074. More Gale Encyclopedia of Nursing and Allied Health Nitroglycerin topical NITROGLYCERIN (nye troe GLI ser in) is a type of vasodilator.

1075. More Gale Encyclopedia of Medicine Nitroglycerin NITROGLYCERIN (nye troe GLI ser in) is a type of vasodilator.

1076. A surgeon inserts an endosc More HLCMS Nitroglycerin NITROGLYCERIN (nye troe GLI ser in) is a type of vasodilator.

1077. Infliximab injection is used to treat adults with rheumatoid arthritis and ankylosing spondylitis, which is a type of arthritis that affects the joints in the spine.

1078. This medicine is also used to treat psoriatic arthritis, which is a type of arthritis that causes pain and swelling in the joints along with patches of scaly skin on some areas of the body.

1079. Naproxen also helps relieve symptoms of ankylosing spondylitis, which is a type of arthritis that affects the joints in the spine.

1080. It is also used to treat psoriatic arthritis, which is a type of arthritis that causes pain and swelling in the joints along with patches of scaly skin on some areas of the body.

1081. It is used to treat adults with psoriatic arthritis, which is a type of arthritis that causes pain and swelling of the joints and patches of scaly skin on some areas of the body.

1082. Some types of STEC frequently cause severe disease, including bloody diarrhea and hemolytic uremic syndrome (HUS), which is a type of kidney failure.

1083. Some types of STEC frequently cause severe disease, including bloody diarrhea and hemolytic uremic syndrome (HUS) which is a type of kidney failure.

1084. Among ill persons, 50% reported being hospitalized, and none reported hemolytic uremic syndrome, which is a type of kidney failure that is associated with E. coli O157:H7 infections.

1085. Cascara is a type of laxative called a stimulant laxative.

1086. Cascara is a type of laxative that might also decrease potassium in the body.

1087. Lymphoma is a type of cancer that affects the lymphatic system, which is part of the bodys immune system.

1088. Lymphoma is a type of cancer that affects the lymph system, which is part of the bodys immune system.

1089. There is a type of headache called Cluseter Headache you may want to discuss with your doctor.

1090. There is a type of headache called Primary Exertional Headache that occurs during or right after exercise.

1091. There is a type of headache called Primary Exertional Headache (PEH) that about 10% of the population gets, mostly men and can be triggered by any type of exercise.