DETECTION OF BURSTY AND EMERGING TRENDS TOWARDS IDENTIFICATION OF RESEARCHERS AT THE EARLY STAGE OF TRENDS

by

SHERON LEVAR DECKER

(Under the direction of Budak Arpinar)

ABSTRACT

Detection of trends is important in a variety of areas. Scientific research is no exception. While several methods have been proposed for trend detection, we argue that there is value on using semantics-based techniques. In particular, we demonstrate the value of using a taxonomy of topics together with data extraction to create a dataset relating publications to topics in the taxonomy. Compared to other approaches, our method does not have to process the content of the publications. Instead, it uses metadata elements such as keywords and abstracts. Using such dataset, we show that a semantics-based approach can detect "bursty" and "emerging" research topic trends. Additionally, our method identifies researchers involved at the early stage of trends. We use known lists of recognized and prolific authors to validate that many of the researchers identified at the early stage of trends have indeed been recognized for their contributions on important research trends.

INDEX WORDS: Trend Detection, Emerging Trends, Bursty Trends, Taxonomy of Computer Science Topics, DBLP, Data Extraction

# DETECTION OF BURSTY AND EMERGING TRENDS TOWARDS IDENTIFICATION OF RESEARCHERS AT THE EARLY STAGE OF TRENDS

by

SHERON LEVAR DECKER

B.S., South Carolina State University 2003

A Thesis Submitted to the Graduate Faculty of The University of Georgia in Partial Fulfillment of the Requirements for the Degree

MASTER OF SCIENCE

ATHENS, GEORGIA

2007

DETECTION OF BURSTY AND EMERGING TRENDS TOWARDS IDENTIFICATION OF RESEARCHERS AT THE EARLY STAGE OF TRENDS

by

SHERON LEVAR DECKER

Major Professor: Budak Arpinar

Committee: John A. Miller
David Himmelsbach

Electronic Version Approved:

Maureen Grasso
Dean of the Graduate School
The University of Georgia
August 2007

# DEDICATION

This paper is dedicated to a father figure in my life, Tony Edmund, my best friend, Kenyatta Wilson, my girlfriend, Carla Phillips, and last but not least my family

Mr. & Mrs. Woodrow Hemingway (Dad & Mom)

Rolando Decker (Brother)

Ellirose Hemingway (Sister)

To you all, I thank you for believing in me from the beginning.

# ACKNOWLEDGEMENTS

I would first like to acknowledge GOD for giving me the gift of life and having someone to turn to when I had no direction and felt I could not continue.

I would like to thank Boanerges Aleman-Meza for your irreplaceable guidance and leadership throughout my entire endeavor in completing my research. Thanks for everything because it was not required of you to help me but you did from the kindness of your heart and I will never forget that. I would like to thank Delroy Cameron for your daily commitment to accomplish whatever had to be completed. Your dedication is something I have never seen in an individual and it has motivated me to sacrifice in order to put myself in a better position to succeed in life. Thank you both for being great friends throughout this journey of life. I will miss you guys.

I would like to thank Dr. Himmelsbach for allowing me the opportunity to pursue a MS degree with the support from ARS/USDA. While working as an intern, you saw that I was more than capable of taking the next step in receiving a higher education and advised me to make that decision. I had no plans after undergrad but you took me under you wing and lead me to where I am now. For that I thank you and will never forget you. You are a great inspiration.

Lastly, I would like to thank Dr. Budak Arpinar and Dr. John Miller. Thank both of you for assistance, direction, and support. Thank you for teaching me what it takes to become a successful computer scientist and for assigning all the hard work in class because it is definitely going to pay off. Thank you both for doing your job and doing it well.

TABLE OF CONTENTS

Page

# LIST OF TABLES

Page

# LIST OF FIGURES

Page

# CHAPTER 1

# INTRODUCTION

One way to keep up with the landscape of research in a field of study is to stay informed with the trends that are occurring in the area. Having knowledge of past, current and emerging trends is quite valuable. For example, a scientist might want to do research in an area that has not been touched on heavily. It can also be valuable in the sense of a businessperson trying to evaluate the risks of investing in a new business. Trend detection has already been applied with the use of text documents, blogs and emails [12, 16, 17, 27]. The detection of trends could be very important for funding agencies such as the National Science Foundation (NSF) in order to determine or justify whether projects in new areas of research are to be funded. Identifying influential researchers on topics could help validate that their funding within certain areas has had a positive/productive impact in the research community. Automated approaches have been built for identifying funding agencies in acknowledgments section of papers [6]. From the standpoint of identifying past trends, one could determine if there is any correlation to the amount of funding provided to a previous area of interest with respect to its success or impact. Identifying participants at the emerging stage of a trend is of importance because it will determine who were the influential people that aided or started the popularity of a given trend. For example, the Association for Computing Machinery (ACM) program recognizes and honors individuals for their achievements in the computer science and information technology fields. The identification of researchers that are identified as “trend setters” could help in determining

the individuals to consider for such awards. The goal of our work is to develop an approach that will detect two types of trends. The first are "bursty" trends, which have the characteristic of having one or more intense periods of activity. Second are "emerging" trends, which are characterized by having an increasing activity over a period of time but not necessarily with a "bursty" behavior.

Gruhl et al. studied detection of "bursts" in the blogsphere for cases where the total number of blog entries on a particular topic exceeded a formulated threshold [12]. They also examined whether these topics could be "de-spiked" to identify an underlying, probably unknown reason for the burst. We use these same ideas, but with different approaches. First, we focus on identifying research topics using different data, namely metadata of publications such as keywords and abstracts. Second, we demonstrate that trends in research can also have "bursts," which we identify based on the total number of publications written on the topic using time intervals of days, months or years. Other work for trend detection in publication data has only managed to use years as the unit of time. Thirdly, we implement "de-spiking" on a research topic to identify other topics that might be the cause of the bursty behavior. This allows analysis to determine if other topics had any impact on the burst (if indeed there was a burst in the total number of publications on the topic). For example, how much of a contribution has the topic "PageRank" had towards the trend in the topic "Ranking"? We also focus on determining who were the authors that published on the research areas at the early stage of the trend. This approach builds upon the work of Gruhl et al. [12], who adopted a simple set of predicates on topics that would allow them to associate particular blog posts appearing at different parts of a topic life cycle.

In other work [12, 17, 27], evaluation of bursts was accomplished using the construction of time graphs, whereas [16] took the approach of using a weighted automaton model. In the context of blogs, posts have specific timestamps associated with themselves to identify when they were created in order to create time sequential graphs. Similarly, emails can be tracked based on arrival structure. In our work, we evaluate bursts in a research area using the time graph approach. In order to construct a graph for a research area we first have to create a dataset that relates papers to one or more research topics. Because Digital Bibliography Library Project (www.informatik.uni-trier.de/~ley/db/) is one of the largest websites that lists computer science bibliography, we decided to use it as a/the data extraction source. We demonstrate how this type of dataset can be created with focused crawling and off-the-self techniques for term extraction (e.g., Yahoo! Term Extraction (developer.yahoo.com)). Extracting data relating topics to publications from DBLP is extremely challenging because DBLP does not contain information relating publications to research areas or topics. We developed methods that create such paper to topic relationships. A publication can then be explicitly related to one or more topics. One of our goals is in demonstrating how this is possible without having to process the content of documents, which in this case, are publications that exist in a variety of document formats (e.g., PDF, PostScript, and HTML). Similar datasets can be created for other research fields such as chemistry or biology. For the purposes of this paper, our approach is tested using a dataset that is focused on research areas of Database, Information Retrieval, Web and Semantic Web, AI and Data Mining. This dataset consists of 78K publications and 40K relationships connecting publications to topics in a taxonomy of Computer Science research areas.

The contributions of our work are two-fold. First, we describe a methodology for building a dataset that contains relationships from publications to topics in a taxonomy of topics. The benefits of this type of dataset is that the papers to topics relationships connect topics in the taxonomy to publications in an existing ontology of publications that was created using DBLP data. Second, we demonstrate a semantics-based approach for determining “bursty” and “emerging” research topic trends together with the capability of identifying researchers at the emerging stages of research areas.

# CHAPTER 2

# BACKGROUND

## 2.1 Ontologies and Taxonomies

As the existing Web evolves into a Semantic Web, the necessity for ontologies is inevitable. Gruber defines an ontology as a specification of a representational vocabulary for a shared domain of discourse – definitions of classes, relations, functions, and other objects [11]. It is used to represent concepts and the relationships between those concepts. Concepts are described with the use of classes, which may contain individuals, molecules, other classes, or a combination of the three. Ontologies are frequently used among applications, researchers, and databases for the purposes of sharing domain knowledge. Hence, sharing of ontologies with others enables integration in other domains so others do not have to develop ontologies from scratch and also reuse of domain knowledge.

Taxonomy is the science of classification, or categorization, of things based on a predetermined system. It is a conceptual framework for analysis of cognitive activities as they actually unfold in a complex work situation. It is intended to be a vehicle for generalization of results of field studies in various domains so as to make it possible to transfer results among domains and to serve needs of research in general in complex work environments [21]. Taxonomies are being vastly used in the fields of computer science, biology, chemistry, etc.

## 2.2 RDF and RDFS (Ontology Language)

Ontologies are developed using description languages. One of the more commonly used languages to encode an ontology is Resource Description Framework (RDF). RDF is an XML application that allows for the denotation of facts and schemata in a web-compatible format, building on an elaborate object-model for describing concepts and relations [25]. It is a foundation that allows encoding, exchanging and reuse of structured metadata. RDF's metadata model is based on the idea that resources can be described with expressions in the form of subject-predicate-object, called triples. Properties of RDF can be used to represent relationships amongst resources. However, the RDF model does not provide any means for declaring these properties or defining the relationships between these properties and other resources. For these reasons, RDF Schema (RDFS) is used with the intent to structure RDF resources. A schema defines not only the properties of the resource (Title, Author, Subject, Size, Color, etc.) but may also define the kind of resources being described (books, Web pages, people, companies, etc) [5].

## 2.3 Semantic Analytics

Semantic analytics is the use of ontologies to analyze content in web resources. This field of research combines text analytics and semantic web technologies like RDF. Systems are becoming exceptionally beneficial that go beyond basic search and integration capabilities by offering users an interface for performing ontological computation and formulating complex

relationship type queries [23]. Querying and inference techniques are two of the many steps typically involved in the process for development of a Semantic Web application. They are needed as a foundation for more complex data processing and enabling semantic analytics and discovery [2].

### 2.4 Semantic Associations

Semantic associations are relevant and meaningful complex relationships between, events, entities and concepts. They provide new and possibly unexpected insights and lend meaning to information, making it actionable and understandable [24]. Semantic Associations can span across multiple domains and may involve any number of intermediate entities and relations [13]. RDF is being widely used for its capabilities of capturing meaning between resources based on how they relate to other resources through such semantic associations.

### 2.5 Emerging Trend

An emerging trend is a topic area for which one can trace the growth of interest and utility over time [9]. Studies have been put forth in order to elucidate new and emerging trends from the empirical, technological, and theoretical perspectives [22]. Being aware of emerging trends is of noteworthy importance for business owners in order to make an effort to predict what the consumer is likely to be in demand for. As the amount of digital information increases, more

and more automated systems are coming into use to aid in the detection of emerging trends for human experts in this area of study.

# CHAPTER 3

# METHOD FOR BUILDING A PUBLICATIONS-TO-TOPICS DATASET

## 3.1 Using DBLP Bibliography Data

In the work of Tho et al., the majority of their dataset of scientific publications was retrieved from websites of academic institutions [27]. We argue that better results are possible when using larger datasets. DBLP is an excellent site that lists bibliography data of more than 885,000 computer science publications. Hence, it is a good dataset choice to demonstrate our approach. For the purposes of this paper, we used a subset of DBLP data that includes a variety of publications in research areas including Databases, Web, Semantic Web, Data Mining, AI, and Information Retrieval. However, the method of building a publications-to-topics dataset is not tied to these areas. A similar subset of DBLP data was used for finding connected researchers [7]. In a similar way as in such work, we list the conferences, workshops, and journals of the papers composing the subset we used, see Table 1. In fact, the list in Table 1 is a superset of that listed in [7]. The subset (95MB) used in our approach was taken from DBLP data as of May 1st, 2007.

**Table 1: Publication Venues of the Papers Included in the Dataset Used (Subset of DBLP)**

| **Conferences (113)** |
|---|
| AAAI, ADB, ADBIS, ADBT, ADC, ARTDB, BERKELEY, BNCOD, CDB, CEAS, CIDR, CIKM, CISM, CISMOD, COMAD, COODBSE, COOPIS, DAISD, DAGSTUHL, DANTE, DASFAA, DAWAK, DBPL, DBSEC, DDB, DEDUCTIVE, DEXA, DEXAW, DIWEB, DMDW, DMKD, DNIS, DOLAP, DOOD, DPDS, DS, DIS, ECAI, ECWEB, EDBT, EDS, EFDBS, EKAW, ER, ERCIMDL, ESWS, EWDW, FODO, FOIKS, FQAS, FUTURE, GIS, HPTS, IADT, ICDE, ICDM, ICDT, ICOD, ICWS, IDA, IDEAL, IDEAS, IDS, IDW, IFIP, IGIS, IJCAI, IWDM, INCDM, IWMMDBMS, JCDKB, KCAP, KDD, KR, KRDB, LID, MDA, MFDBS, MLDM, MSS, NLDB, OODBS, OOIS, PAKDD, PDP, PKDD, PODS, PPSWR, RIDE, RULES, RTDB, SBBD, SDB, SDB, SDM, SEMWEB, SIGMOD, SSD, SSDBM, TDB, TSDM, UIDIS, VDB, VLDB, W3C, WEBDB, WEBI, WEBNET, WIDM, WISE, WWW, XP, XSYM |
| **Journals (28)** |
| AI, AIM DATAMINE, DB, DEBU, DKE, DPD, EXPERT, IJCIS, INTERNET, IPM, IPL, ISCI, IS, JDM, JIIS, JODS, KAIS, SIGKDD, SIGMOD, TEC, TKDE, TODS, TOIS, VLDB, WS, WWW, WWJ |

We considered various RDF-encoded datasets of DBLP data, namely, the D2RQ-generated RDF data from DBLP [4], Andreas Harth's DBLP dataset in RDF (sw.deri.org/~aharth/2004/07/dblp/), and our own SwetoDblp ontology (lsdis.cs.uga.edu/projects/semdis/swetodblp/). We selected SwetoDblp because it allows the possibility to exploit the benefits of representing and aggregating data in RDF. For example, SwetoDblp includes affiliation data based on heuristics using the homepage information of the authors. Hence, the individuals participating in trends could be listed together with their affiliation. The other side of the coin is that it is possible to determine all the trends in which a given university or organization is associated (through the people affiliated with it). Other research efforts not necessarily related to trend detection have highlighted the value of using semantics for describing data of publications [1, 10].

Table 2 lists the number of instances in the major classes. Compared to SwetoDblp, this subset is around one tenth in terms of number of entities. SwetoDblp is a dataset of 845 MB file size; the subset we used is 95 MB file size. Throughout this paper, we will refer to SwetoDblp whenever a particular aspect of such ontology is highlighted, otherwise we will simply refer to DBLP.

**Table 2: Instances in Main Classes in the Subset We Used Compared to DBLP**

| Main Classes | Subset | DBLP |
|---|---|---|
| Proceeding (of conferences, etc) | 857 | 8,665 |
| Articles in proceedings | 51,202 | 532,758 |
| Articles in journals | 25,973 | 328,792 |
| Authors | 67,366 | 539,301 |

The four most important classes that were used in our subset are proceedings within conferences, authors, articles in proceedings, and articles in journals. Each publication that is part of a proceeding is related to authors and a proceeding with an “inproceedings” and “author” relationship. Determining the proceedings with which a paper is located in is very important in our work for plotting data at different time intervals. Each publication contains the year the paper was published. Although this suffices for creating a time graph that represents the years of research papers, it is not enough information to plot papers by day and/or month. In order to overcome this dilemma, we extracted exact dates from each proceedings title. Our approach is an improvement over other approaches that are limited in plotting data only based on years [7, 27]. For example, the proceeding title “Graphics and Robotics, Dagstuhl Castle, Germany, April 19 –

22, 1993” has an exact date at which such meeting took place. We used methods that extracted these dates from each proceeding in the dataset (if indeed there was a date in the title). With this information we can explicitly relate many publications to an exact date, namely, the date at which the paper was presented in its corresponding conference, workshop, or symposium.

Out of the all the proceedings in our dataset, we were able to extract dates for 94% of the proceedings. For the papers that were not able to get associated to an exact date, a check was done to see if the year of the paper matched the year of the last-modified-date (metadata value in DBLP) for that paper and if so then such exact date was used.

### 3.2 Taxonomy of Topics

There are computer science classification systems readily available that could have been re-used in our approach. For instance, ACM's Computing Classification System (CCS) (acm.org/class/1998/) provides a categorization of computer science related topics intended to reflect the current state of the field. It contains eleven primary research areas each including numerous subtopics. However, the system is comprised of a very “broad” four-level tree of topics that would not be very beneficial recognizing topics that are manifesting today. For example, a publication entitled “Semantic analytics on social networks: experiences in addressing the problem of conflict of interest detection” was classified with ACM's CCS with the primary topic 'Information Systems' because no other topics such as social networks, semantic analytics, or conflict of interest were available. Therefore, we developed our own taxonomy of computer science topics that would help identify “newer” terms. Identification of newer terms is

advantageous for the purpose of recognizing possible emerging trends that might be included in a taxonomy of topics.

Building a taxonomy of research topics is a significant endeavor. In order to give structure to the research topics a taxonomy was created that contains Computer Science topics. The taxonomy of topics has very good coverage for the areas of Databases, Web, Semantic Web, AI, Information Retrieval and Data Mining. Other topics in computer science are also included but at lesser depth (e.g., Computer Architecture). We adapted our taxonomy of topics to that of CoMMA ontology, which has over 420 concepts "arranged in a taxonomy with a maximal depth of 12 levels, more than 50 relationships and more than 630 terms to label these primitives" [8]. We also verified and adjusted the organization of the topics of the taxonomy based on the AKT ontology [20].

The structure of our taxonomy was put together by determining how close topics are related. Our approach began by first retrieving all the URLs of the publications of each research topic term within our dataset from which the terms were included within. We then added each URL into a *set* for each term. Relationships among terms were identified using measures calculated from the intersection of the sets of two terms divided by the union of the sets. This would produce a measure ranging from 0 (which implies the two topics are not related) to 1. Pairs of terms with a value above 0.05 were considered to be related terms. The identification of relationships aids in building a tree-level organization of topics that can later turn into a taxonomy. Figure 1 illustrates examples of topics and their identified relations. Other approaches have done similar work in identifying relationships of topics. In the work by Mika [18], research topics were identified specifically from the interests of researchers within a Semantic Web

community. The associations between the topics were based on the number of researchers who have an interest in the given pair of topics. Our approach instead identifies computer science topics by means of crawling of the DBLP dataset and further data extraction; whereas in their work the topics were already known based on the supplied interests of researchers from FOAF. The work of Velardi [28] is an example of research on taxonomy learning. In our work, we intend to demonstrate that the basic steps for suggesting terms in building a taxonomy can be achieved with off-the shelf tools such as Yahoo! Term Extraction.

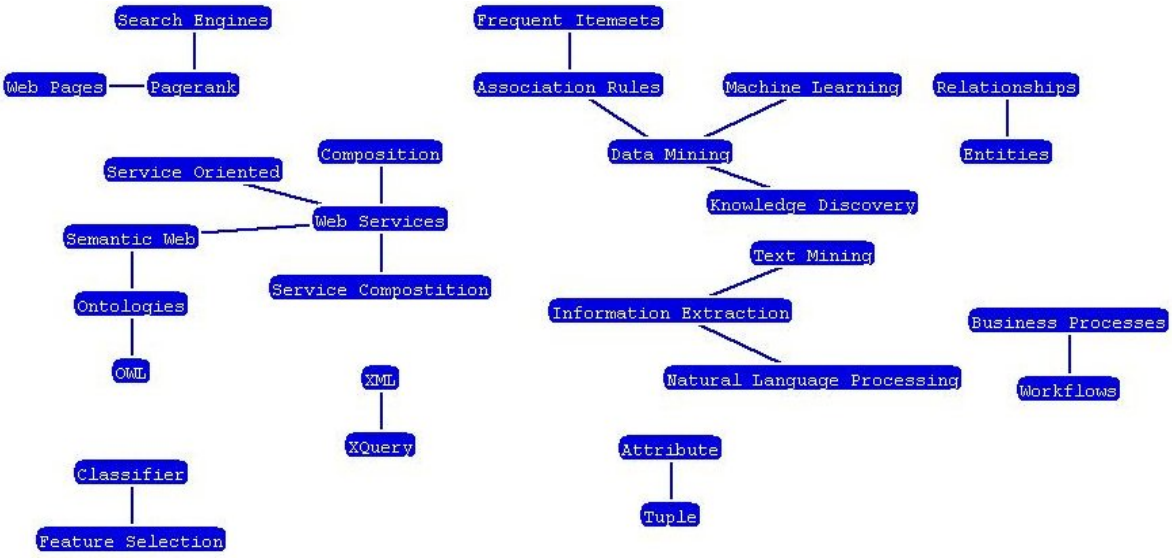


**Figure 1: Snippet of Identified Relationships among Terms**

We created our taxonomy taking into account lessons learned from an earlier effort on creating a small ontology of topics in Semantic-Web

(lsdis.cs.uga.edu/projects/semdis/iswcdemo2006/). Our taxonomy is comprised of 344 research topics from research areas and over 200 synonyms thereof. The taxonomy is available online (http://cs.uga.edu/~cameron/swtopics/taxonomy).

### 3.3 Paper to Topics Relationship

The information in DBLP is not sufficient to determine research topics of publications. For this reason, we developed methods to create paper-to-topic relationships. Creating these relationships was not a straightforward process (refer to Figure 1). The key aspect of our method is how we use the electronic edition “ee” URL literal value of individual papers (metadata value in DBLP) to retrieve additional information of publications. Based on such URL, we performed focused crawling for URLs having doi.acm.org, doi.ieeecomputersociety.org, or dx.doi.org/10.1016 prefixes.

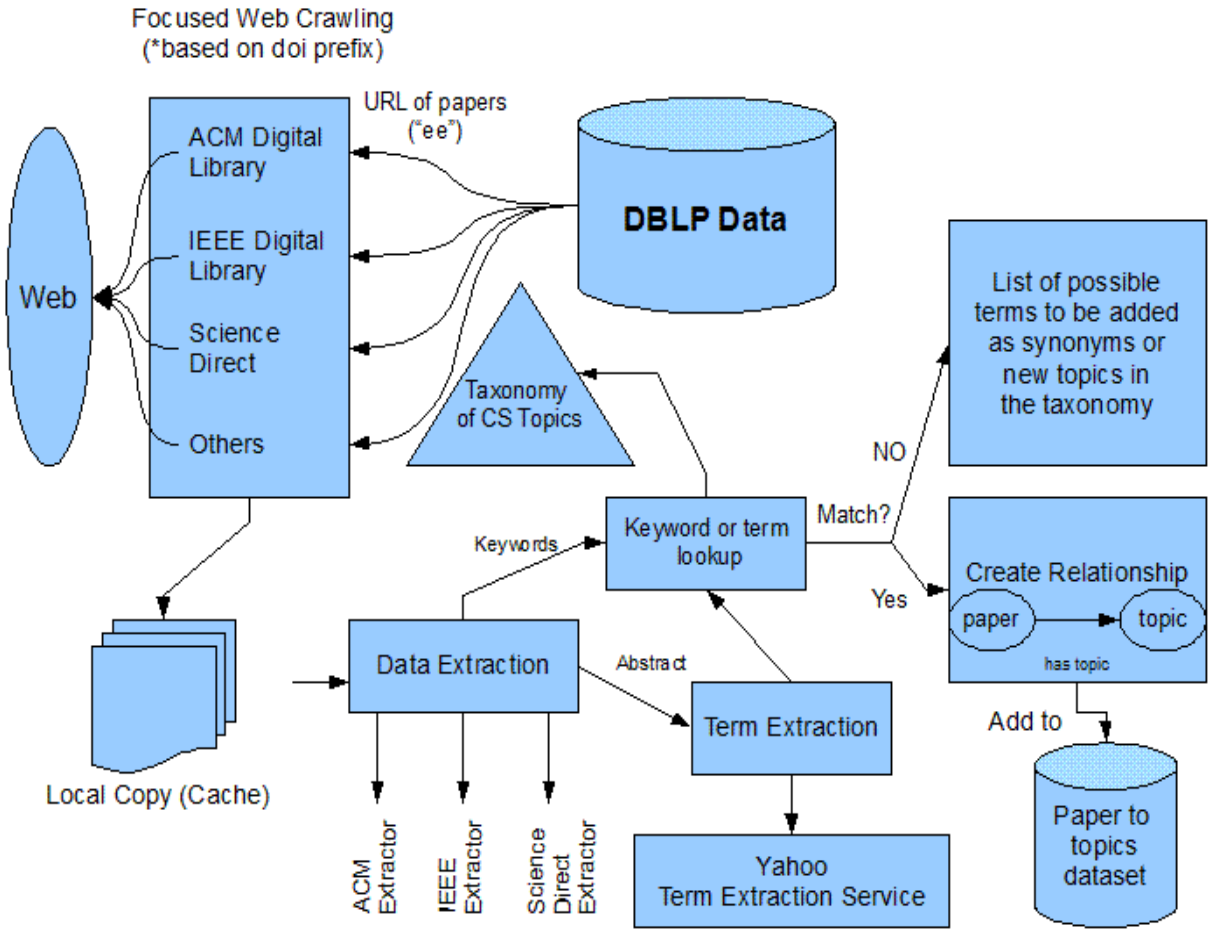


**Figure 2: Overview of Creation of Papers-to-Topics Relationships**

Some publishers' sites, such as Springer, were difficult for document extraction. If we could have extracted data from such large publisher site, then it would very likely improve the quality of our results. Crawled pages were stored in a local cache, from which data extraction

methods obtained keywords and abstracts (when available) for the purpose of identifying a surplus of terms that are related to each publication.

Exploiting structure in different types of web content requires nontrivial data capture tasks [26]. In our case, the data scrapping we performed has the benefit that once data is extracted about a publication, such data is not expected to change. For example, the listed keywords and the abstract of a journal article will always be the same. We experimented using metadata of keywords and abstracts separately. Using keywords alone brings limited data that does not have much added value from the research areas included in ACM's Computing Classification System. On the other hand, by incorporating terms extracted from the abstracts the method aided in identifying “newer” terms. In order to identify the key terms in abstracts, we used the Yahoo! TermExtraction API to determine, based on the input text, what are the most significant words or phrases. Each extracted term, keyword, and phrase of a paper was looked up in the taxonomy of topics to find matches with the name of research topics (or synonyms). If there was a match, a relationship from the paper to the research topic in the taxonomy was established. Otherwise, the terms were kept for possible consideration in identifying “newer” terms to aid in improving the overall taxonomy. The terms were also used as metadata for each publication for the purposes of identifying publications based on a keyword search. We define newer terms as terms that have not appeared within publications before a certain year, in this case we selected the year 2005. Table 3 lists examples of terms that best illustrate newer terms identified with our approach. This was accomplished by determining which papers within our dataset labeled each term as keywords or included the term within its abstract and then retrieved the dates of those publications. A benefit of this approach is that it can keep up with changes in

the field. In fact, Hepp [14] pointed out the need for ontology engineering methods to quickly reflect domain changes to keep ontologies up to update.

**Table 3: Some of the Identified Terms Appearing on Year 2005 and Afterwards**

| |
|---|
| Friendship, grid middleware, grid technology, phishing, protein structures, service oriented architecture (SOA), social network analysis, spam, wikipedia |

In the case of keywords of a paper, the process was similar but without need of term extraction. Two more methods were used. The first consisted of using the names of sessions in conferences as keywords for papers in such sessions. The second is a heuristics that assigns topics to all papers in a conference series, but this is only applicable for very specialized conferences. At the end, 40,718 total relationships from paper to topics were determined. Table 4 lists a summary of how many such relationships were extracted from each site and by using keywords alone.

**Table 4: Total Number of Paper to Topic Relationships Created From Extraction**

| Data Source and/or Data Extraction Method | Relationships (Paper to Topic) | Papers With Relationships to Topics in Taxonomy |
|---|---|---|
| ACM (Keywords) | 2,795 | 1,859 |
| Science Direct (Keywords) | 780 | 631 |
| IEEE (Keywords) | 617 | 454 |
| ACM (Abstract/Terms Extraction) | 5,641 | 3,574 |
| Science Direct (Abstract/Terms) | 2,330 | 1,688 |
| IEEE (Abstract/Terms) | 2,850 | 1,786 |
| Crawling (Session-Names) | 476 | 473 |
| Conference Topics (Heuristics) | 25,229 | 23,083 |

As a means to determine which were the most common terms accumulated, we kept a record of how many times each term appeared. This allowed us to identify terms and phrases that were highly used as keywords and words within abstracts. Table 5 lists ten of the most frequently identified terms within the last ten years.

**Table 5: Few of the Top Terms Identified From URL Extraction within Last Ten Years**

| Topic | 1998 | 1999 | 2000 | 2001 | 2002 | 2003 | 2004 | 2005 | 2006 | 2007 |
|---|---|---|---|---|---|---|---|---|---|---|
| Algorithm(s) | 87 | 99 | 111 | 89 | 219 | 222 | 381 | 418 | 608 | 71 |
| Classifier(s) | 0 | 7 | 1 | 2 | 33 | 30 | 47 | 80 | 94 | 5 |
| Data Mining | 12 | 10 | 20 | 13 | 46 | 62 | 88 | 104 | 184 | 8 |
| Databases | 13 | 17 | 19 | 19 | 28 | 32 | 43 | 53 | 63 | 6 |
| Semantic Web | 0 | 0 | 0 | 4 | 13 | 24 | 102 | 85 | 96 | 14 |
| Semantics | 19 | 16 | 26 | 22 | 28 | 24 | 90 | 75 | 86 | 11 |
| Web Service(s) | 0 | 0 | 0 | 0 | 4 | 2 | 67 | 82 | 69 | 1 |
| XML | 0 | 4 | 4 | 11 | 22 | 20 | 36 | 58 | 54 | 1 |

In identifying some of the most frequently identified terms in our approach, we were able to make three key observations pertaining to the results. First, we noticed that terms can be covered in a wide arrange of areas. Therefore, this may constitute for an extremely high volume count of a term compared to other terms. For example, the term *Algorithms* is a very broad term that is not only used as a reference to a research area but also as a means of defining or describing a particular method or technique. This is probably the reason why it appears so many times. Secondly, for a term such as *Databases*, which one would expect to appear more times than shown, we discovered that the total number of appearances is relatively small due to the large amount of synonyms used to represent this particular term. For example, data base, data bases, database management system, database management systems, and DBMS. Hence, if *Databases* is a topic in a taxonomy, then its synonyms should be added as alternate spellings of the term. Thirdly, we were able to identify broader terms, such as the term *Semantics*, which has been used in literature for several years. Although this term has been long used, we were able to detect related terms that have emerged within recent years, case in point being the term *Semantic Web*. This shows that the total number of appearances for these broader terms could be due to newer terms that are related to terms that have been used for a longer time.

# CHAPTER 4

## DETECTION OF TRENDS USING BIBLIOGRAPHY AND TOPICS DATA

Our method is able to detect two types of trends: bursty trends and emerging trends. In addition, it is possible to identify researchers at the emerging stages of a research topic. Figure 3 provides an overview of our approach. Several steps are taken in order to (1) retrieve all the information pertaining to a research area; (2) determining if a research topic is a bursty and/or emerging trend.

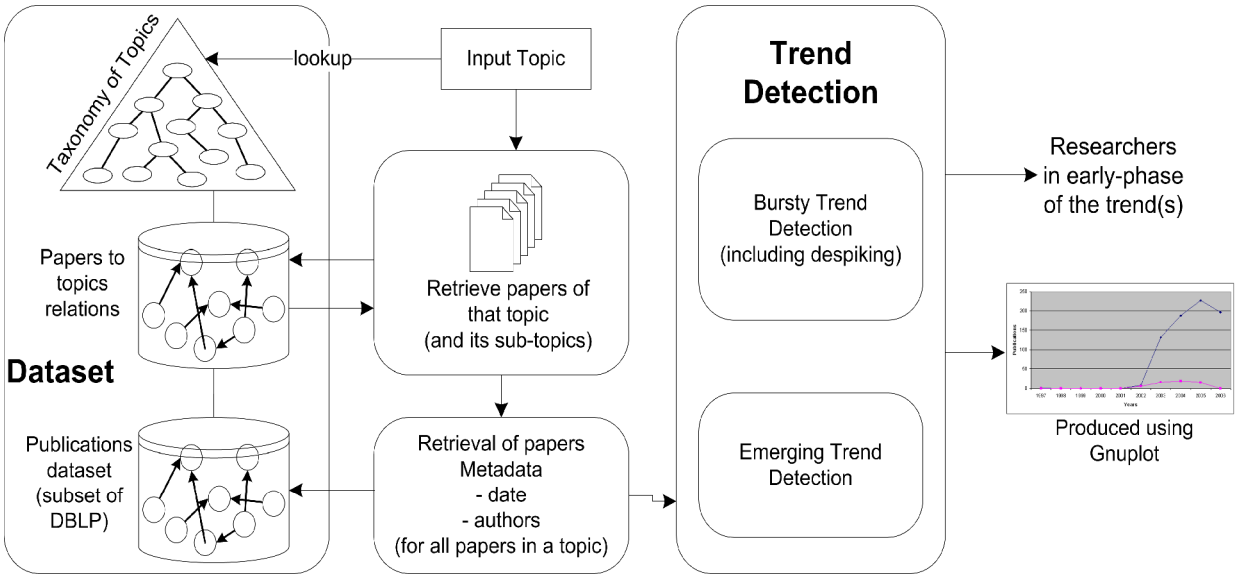


**Figure 3: Overview of Bursty and Emerging Trend Detection and Researcher Identification**

The information gathered on a research topic through our approach is very critical in our trend detection process. The first step is determining whether a given research topic is in the taxonomy. Second, the publications that are associated with the topic (and its sub-topics) are retrieved from our DBLP publications subset. The publications could not have been identified without the paper-to-topics dataset. Then, metadata such as authors and dates are used in detecting whether a research topic is a trend. The benefit of the taxonomy is that all subtopics of a topic are considered. Moreover, publications that are associated with the topic based on the keyword metadata could also be used as an approach for detection of publications relating to the specified topic. This approach is beneficial when trying to identify "newer" topics that are not presented within the taxonomy.

### 4.1 Detection of Bursty Trends

Formulas for four predicates, which were devised in order to classify individuals to a region within a time graph of blog posts where they posted the most, were adapted from the work done by Gruhl et al. [12]. We use a similar formula of one of the predicates for determining if a research topic is a bursty trend. Figure 4 is graph of an actual topic that illustrates how bursty trends are detected.

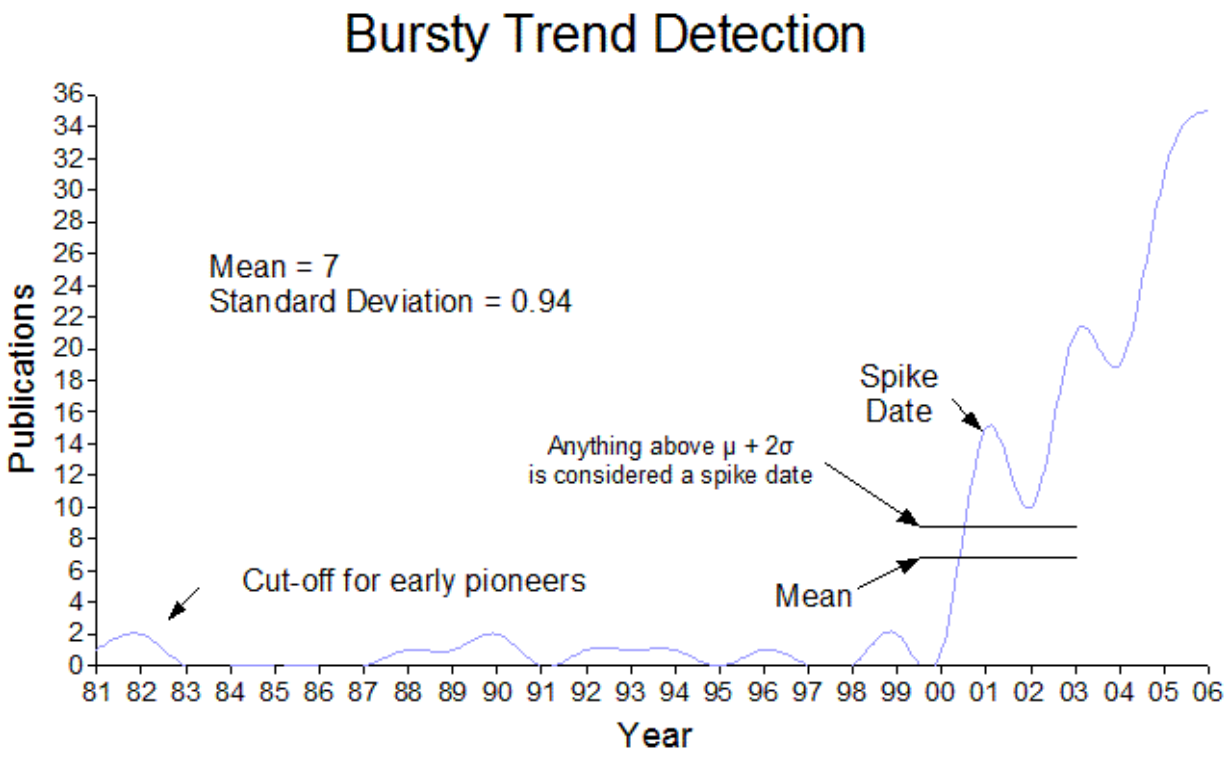


**Figure 4: Bursty Trend Detection Overview**

If the total number of publications for a particular research topic is greater than a threshold value ($\mu + 2\sigma$) for any day, month or year (depending on the time unit being used), the research topic is considered to be a bursty trend. Figures 5 and 6 show examples of bursty trends that were detected by our approach. These are "Data Model" and "Semantics." Interestingly, both have had increased popularity in the last few years and both have also appeared in the literature over the last 30 years.

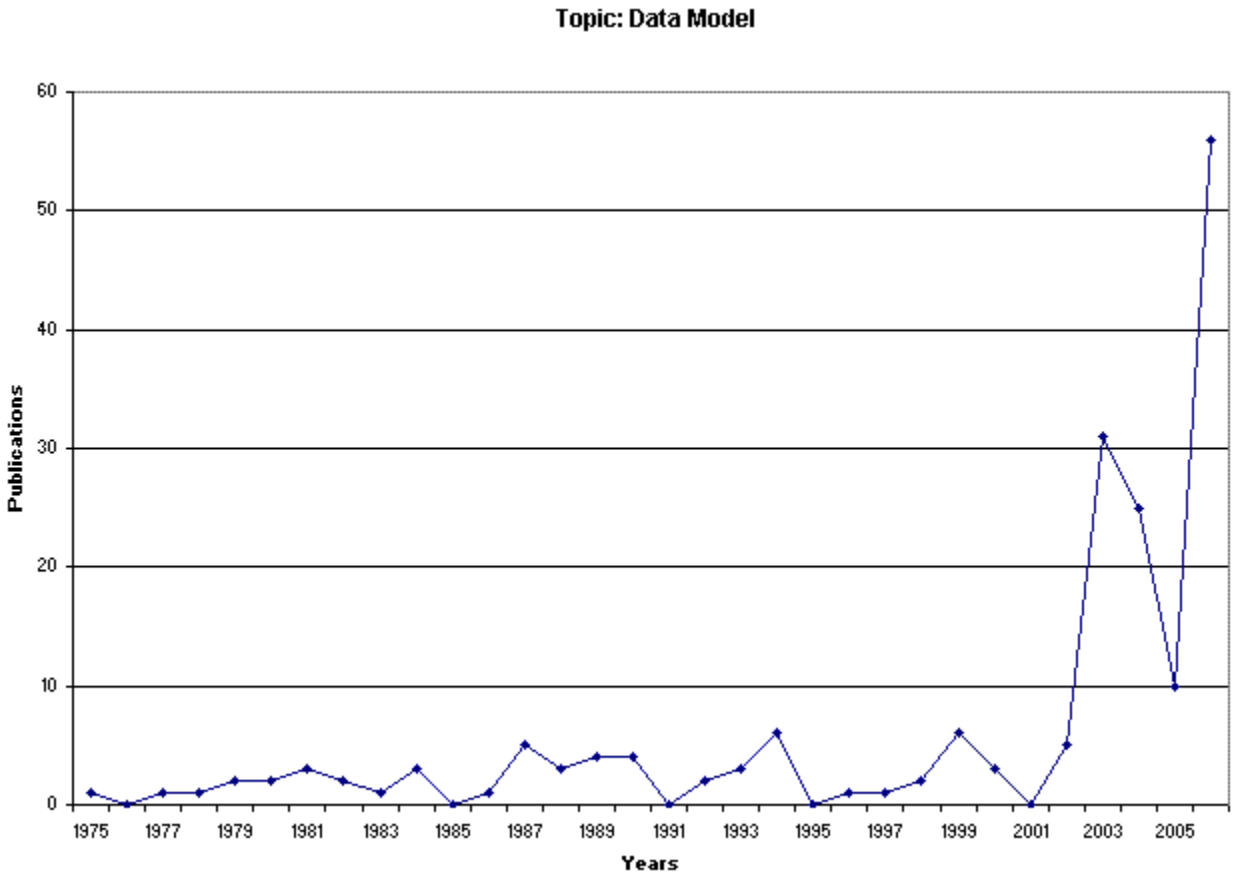


**Figure 5: Example of Bursty Trend for Topic: Data Model**

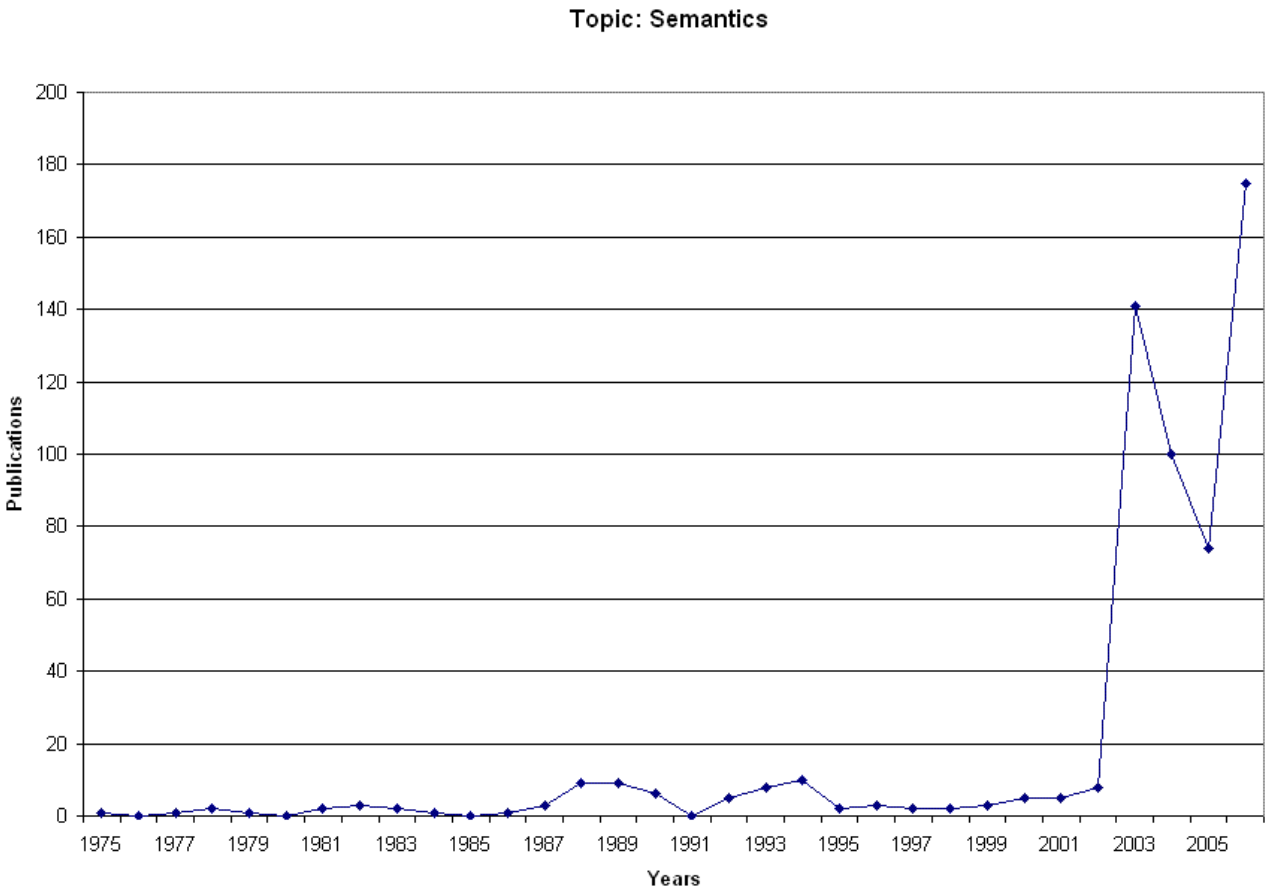


**Figure 6: Example of Bursty Trend for Topic: Semantics**

### 4.2 Detection of Emerging Trends

A method for detecting emerging trends used statistical information of documents published in research areas [27]. We implemented their algorithm for identification of emergent trends to apply it with our dataset. Their method determines whether there has been a significant increase in the total number of publications within recent years. Emerging trends do not

necessarily exhibit a bursty behaviour. Figures 7 and 8 show examples of emerging trends that were detected with our approach. We purposely excluded the current year (2007) from our data for the reason of it not being a complete year as of yet.

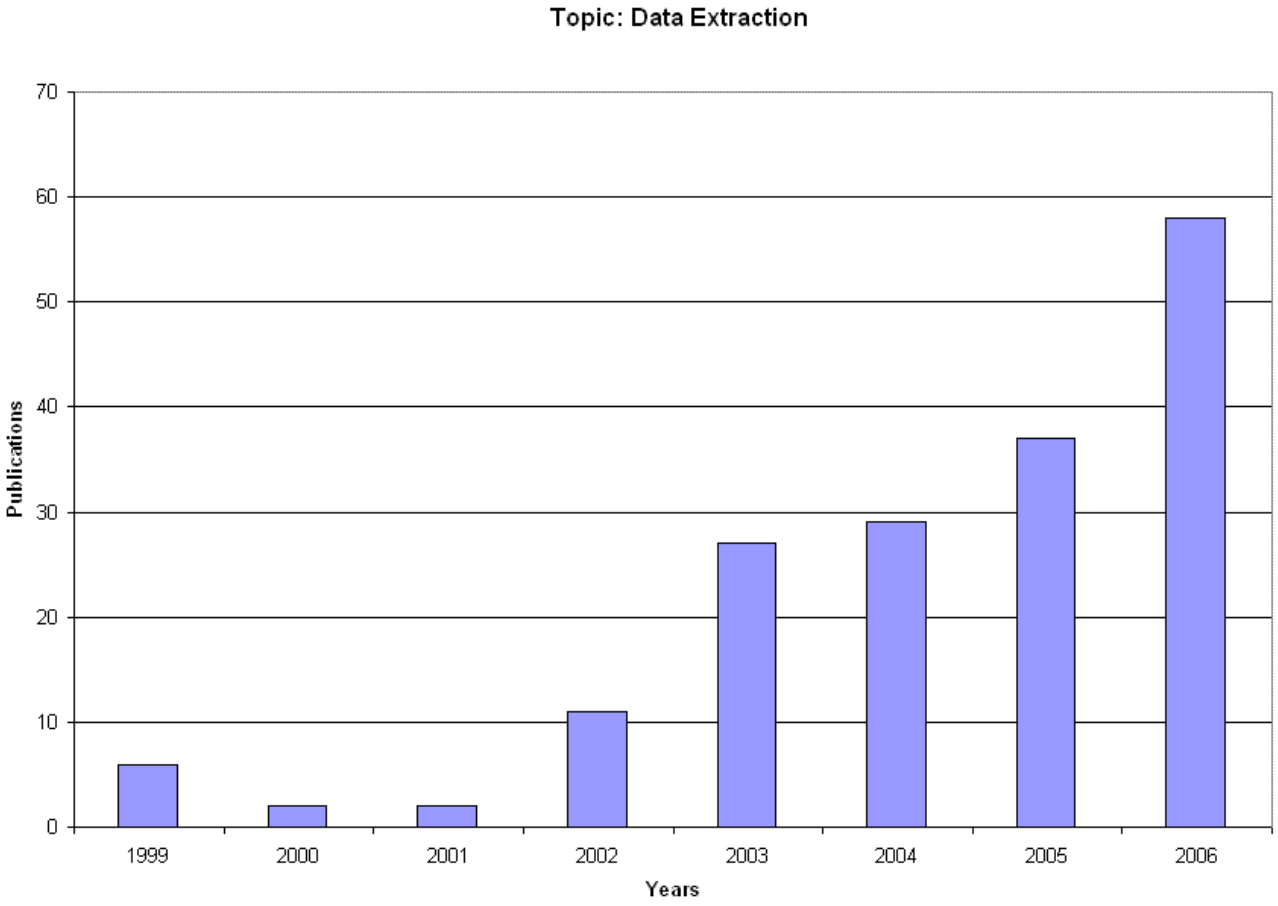


**Figure 7: Example of Emerging Trend for Topic: Data Extraction**

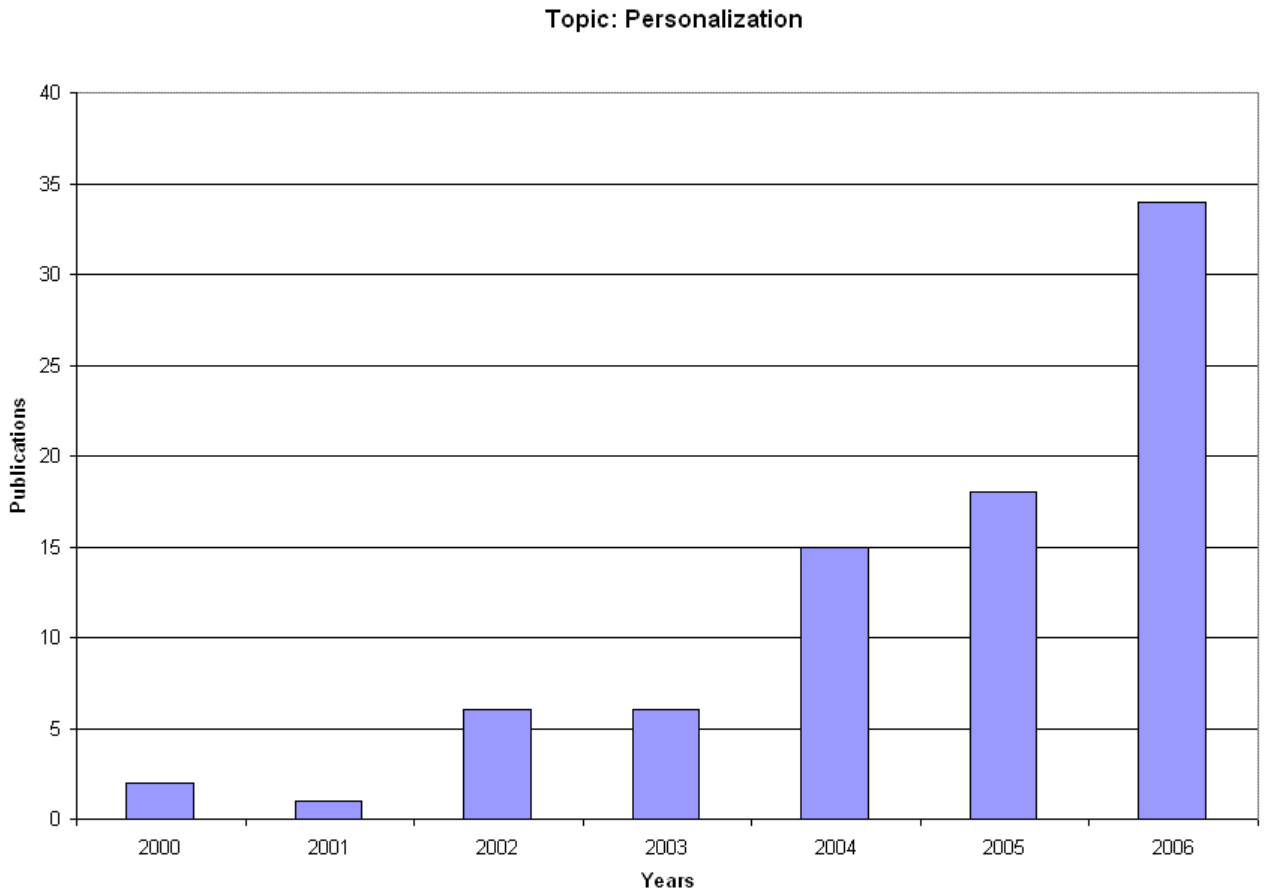


**Figure 8: Example of Emerging Trend for Topic: Personalization**

# CHAPTER 5

# EVALUATIONS AND RESULTS

## 5.1 Influential Researchers

After identification of trends, it is possible to determine the people involved at the early stage of the development of a trend. Al-Sudani et.al [1] described this idea intended for finding knowledgeable personnel in certain areas of interest. However, they used a much smaller dataset of publications. In fact, they point out that data collection/extraction is a time-consuming task. We believe that our approach circumvents such problem by using the metadata itself of publications for selecting URLs that contain keywords and terms metadata to be extracted. Some manual work has to be done, in our case, for the creation of a web-scrapper for a specific web source such as ACM Digital Library. The advantage is that once such metadata is extracted, it can be safely assumed that it is not going to change. That is, the keywords of a published article will always remain unchanged.

Evaluating whether we found influential researchers at the emerging stages of a research topic is challenging. We chose to compare the individuals appearing at the early stage of a trend with respect to six existing lists that contain highly recognized or prolific computer scientists. The lists are as follows: (1) ACM Fellows, (2) IEEE Fellows, (3) DBLP People that are in Wikipedia, (4) H-Index, (5) Prolific Authors, and (6) Centrality Score. (1) The ACM fellows list (fellows.acm.org/) includes members recognized in the Computer Science and Information

Technology for their professional, technical and leadership contributions. (2) The IEEE Fellows list (ieee.org/web/membership/fellows/new_fellows.html) includes an elite group from around the globe. The IEEE looks to the Fellows for guidance and leadership as the world of electrical and electronic technology continues to evolve. An IEEE Fellow shall have contributed importantly to the advancement or application of engineering, science and technology, bringing the realization of significant value to society. (3) Researchers that have a Wikipedia page and also appear in DBLP can be arguably considered as persons that have had some sort of impact or recognition in Computer Science. The content of Wikipedia is compiled from a large number of participants. However, the mechanisms in Wikipedia make it extremely difficult to create (and keep) a new Wikipedia entry for a person. That is, a Wikipedia page about a person can be created only if such person is arguably famous, has an important position, has important achievements, etc. Hence, we assume that Wikipedia entries about Computer Science researchers can be viewed as evidence of their important contributions. We extracted URLs for such persons by processing a dataset of such persons made available by a recent effort on extracting semantics from Wiki content [3]. (4) The h-index is defined as a measure to characterize the scientific output of a researcher, where h is the number of papers with citation number higher or equal to h [15]. We used an existing list of computer science researchers with h-index of 40 or higher (www.cs.ucla.edu/~palsberg/h-number.html). (5) Prolific authors from DBLP data (www.informatik.uni-trier.de/~ley/db/indices/a-tree/prolific/index.html) are individuals that have the most publications within DBLP. (6) There are known methods to identify participants in a network that are highly connected to the rest. The *closeness centrality* measure identifies how close an author is, on average, to all other authors in the network. Authors with low *closeness*

values are connected to many authors within short path distances. Hence, it could be said that their ‘influence’ in the network is high. We computed centrality as the average of the shortest path that an author has to each author. Table 6 lists the top 10 *central* authors from the largest connected component in DBLP-subset. The first column lists authors computed by simply taking their name as they appear in DBLP.

**Table 6: Top 10 *Centrality* Authors in DBLP-Subset**

| **Centrality using Name** | | **Centrality using *same-as* Information** | |
|---|---|---|---|
| **Score** | **Author Name** | **Score** | **Author Name** |
| 4.0578 | Gio WiederHold | 3.9859 | Gio WiederHold |
| 4.1527 | Richard T. Snodgrass | 4.0517 | Umeshwar Dayal |
| 4.1900 | Umeshwar Dayal | 4.0616 | Richard T. Snodgrass |
| 4.2020 | Philip A. Bernstein | 4.0825 | Elisa Bertino |
| 4.2025 | Elisa Bertino | 4.1028 | Christos Faloutsos |
| 4.2087 | Christos Faloutsos | 4.1335 | Philip A. Bernstein |
| 4.2232 | Kenneth A. Ross | 4.1431 | Christian S. Jensen |
| 4.2299 | Hector Garcia-Molina | 4.1487 | Jiawei Han |
| 4.2340 | David Maier | 4.1535 | Kenneth A. Ross |
| 4.2427 | Christian S. Jensen | 4.1605 | Erich J. Neuhold |

It has been noted that DBLP does not have unique ID for authors [7]. However, it could be said that the name of an author plays the role of a primary key. For the cases when different persons have the same name, a numerical value is appended in the name to differentiate the two entries in DBLP. For the cases when the same person is referred to in two (or more) forms, then such names (i.e., aliases) are related explicitly, we refer to these as ‘same-as’. Common reasons for people having two names are the use of a shortened name (e.g., Tim

and Timothy) and changes due to addition of hyphened name or middle initial. There are very few entries in DBLP data for authors with more than one name – probably due to the difficulty of detecting such ambiguities automatically. However, it is quite important to make use of information stating that two names refer to the same person. Otherwise, the publications count of an author that has two names would be incorrect. Similarly, co-authorship measures would miss out due to incorrectly counting the right number of co-authors. We compared results obtained when 'same-as' information is used in computing centrality scores of authors. Table 7 lists a couple of examples of authors that appear in DBLP-subset with more than one name. Each name appears with its own centrality score. It is noticeable how much of a change exists on the computed centrality score in the case of Alon Y. Halevy when both of his names spellings are considered. In the case of Timothy W. Finin, his centrality score is also smaller but his position among all computed centrality scores moves from 94 to 101. This happens because the positions of authors computed using same-as information affect not only authors that have more than one name, but also the scores of other authors in the network. This is quite evident in the second column in Table 7, which lists authors when their centrality score is computed using same-as information. It is interesting that the effect of using same-as information is such that the top *centrality* authors differ in both columns.

**Table 7: Examples of Improved Centrality Score by considering the 'same-as' Information**

| **Using 'same-as' Information** | | **Without 'same-as' Information** | |
|---|---|---|---|
| Name of researcher | Centrality score | Names of researcher in the dataset | Centrality score |
| Alon Y. Halevy (37) | 4.2707 | Alon Y. Levy (51)<br>Alon Y. Halevy (111) | 4.4026<br>4.5498 |
| Timothy W. Finin (101) | 4.4051 | Timothy W. Finin (94)<br>Tim Finin (1430) | 4.5123<br>5.0747 |

We measured the overlap in these six lists and found somewhat little overlapping among the acknowledged people in these lists. Table 8 shows the results of the total number of recognized people who appeared in each list.

**Table 8: Comparing Overlap of Lists of Recognized/Prolific Researchers**

| | **# Individuals Appearing In** | **Percentage of Total** |
|---|---|---|
| 1 List | 4,464 | 86.53% |
| 2 Lists | 577 | 11.18% |
| 3 Lists | 97 | 1.88% |
| 4 Lists | 21 | 0.41% |
| 5 Lists | 0 | 0.00% |
| 6 Lists | 0 | 0.00% |

The individuals detected by our method appearing in the early stage of trends can then be compared to the lists before mentioned. However, before such comparison, a process was executed to exclude researchers that do not necessarily publish a lot based on using a measure of collaboration strength [19]. We found that a threshold of 1.0 was sufficient for excluding authors

that, for example, has just one or two papers. Table 9 shows a comparison of the overlap of the six lists plus the list of all researchers at the early stage of research trends identified by our method. This shows that our method detects many of the recognized/prolific authors. In fact, the relative percentages of both lists are very similar.

**Table 9: Comparing our List with Overlap of Lists of Recognized/Prolific Researchers**

| | **# Individuals Appearing In** | **Percentage of Total** |
|---|---|---|
| 1 List | 5183 | 86.34% |
| 2 Lists | 617 | 10.28% |
| 3 Lists | 168 | 2.8% |
| 4 Lists | 28 | 0.46% |
| 5 Lists | 7 | 0.12% |
| 6 Lists | 0 | 0.00% |
| 7 Lists | 0 | 0.00% |

Table 10 shows an example of researchers detected by our approach in the emerging stages of a research topic. These are cases where there is exact match of a recognition they have been given with respect to the topic where they were identified as possible “trend setters.” The column Contribution in the table contains verbatim text from the corresponding list (either ACM Fellows site or a description from Wikipedia).

**Table 10: Recognized researches from trend detection**

| Topic | Person | Appears in List | Contribution |
|---|---|---|---|
| Association Rules | Rakesh Agrawal | ACM Fellow<br>H-Index<br>Prolific Author (167) | "... contributions to data mining" |
| Database | E.F. Codd | ACM Fellow | "... contributions to the theory and practice of database management systems" |
| Information Extraction | Steve Lawrence | Prolific Author (58)<br>Wikipedia Person | "Among the group ... responsible for the creation of the Search Engine/Digital Library CiteSeer" |
| Knowledge Discovery | Jiawei Han | ACM Fellow<br>H-Index<br>Prolific Author (274) | "For contributions in knowledge discovery and data mining" |
| Artificial Intelligence | Raymond Reiter | ACM Fellow<br>Prolific Author (71) | "... contributions to artificial intelligence..." |
| Data Mining | Ming-Syan Chen | ACM Fellow<br>Prolific Author (172) | "... contributions to query processing and data mining" |
| Information Extraction | C. Lee Giles | ACM Fellow<br>Prolific Author (144) | "... contributions to information processing and web analysis" |
| Knowledge Acquisition | Rudi Studer | Prolific Author (130)<br>Wikipedia Person | "Head of the knowledge management research group at the Institute AIFB" |
| Query Languages | Donald D. Chamberlin | ACM Fellow<br>IEEE Fellow | "For contributions to database query languages" |

There is a close relationship between determining topics of papers used together with their date to find trends in research areas to the use of such topics in determining expertise of authors. The benefits of using semantics for expressing expertise or areas of interest for persons have been highlighted in a variety of scenarios and applications (Aleman-Meza, 2007). In fact, the ExpertFinder Initiative intends to identify use cases, challenges, techniques, etc. for semantics-based representation, retrieval, and processing of expertise data (rdfweb.org/topic/ExpertFinder).

### 5.2 De-spiking

De-spiking is the notion of figuring out whether there was some other topic(s) that substantially contributed towards a burst, which is also called spikes. For example, the topic like "Ranking" can be used to relate to several types on ranking. De-spiking removes highly published subtopics that were used in the statistical information of a primary research topic for the purposes of analyzing what the cause of bursts was in a topic. This is achieved with the same method used for detecting a bursty trend. For each subtopic of a research topic, if it is determined that there is a spike in the total number of publications for a given day, month or year (depending on the time unit), then that subtopic is removed from the primary research topic and then re-plotted. Figures 9 and 10 show topics that were detected as bursty trends and the results after the subtopics were de-spiked. It is interesting to see that PageRank is indeed a topic that substantially contributes to the topic Ranking. In the case of de-spiking the topic Service, the contribution of topic Web Services is even more noticeable.

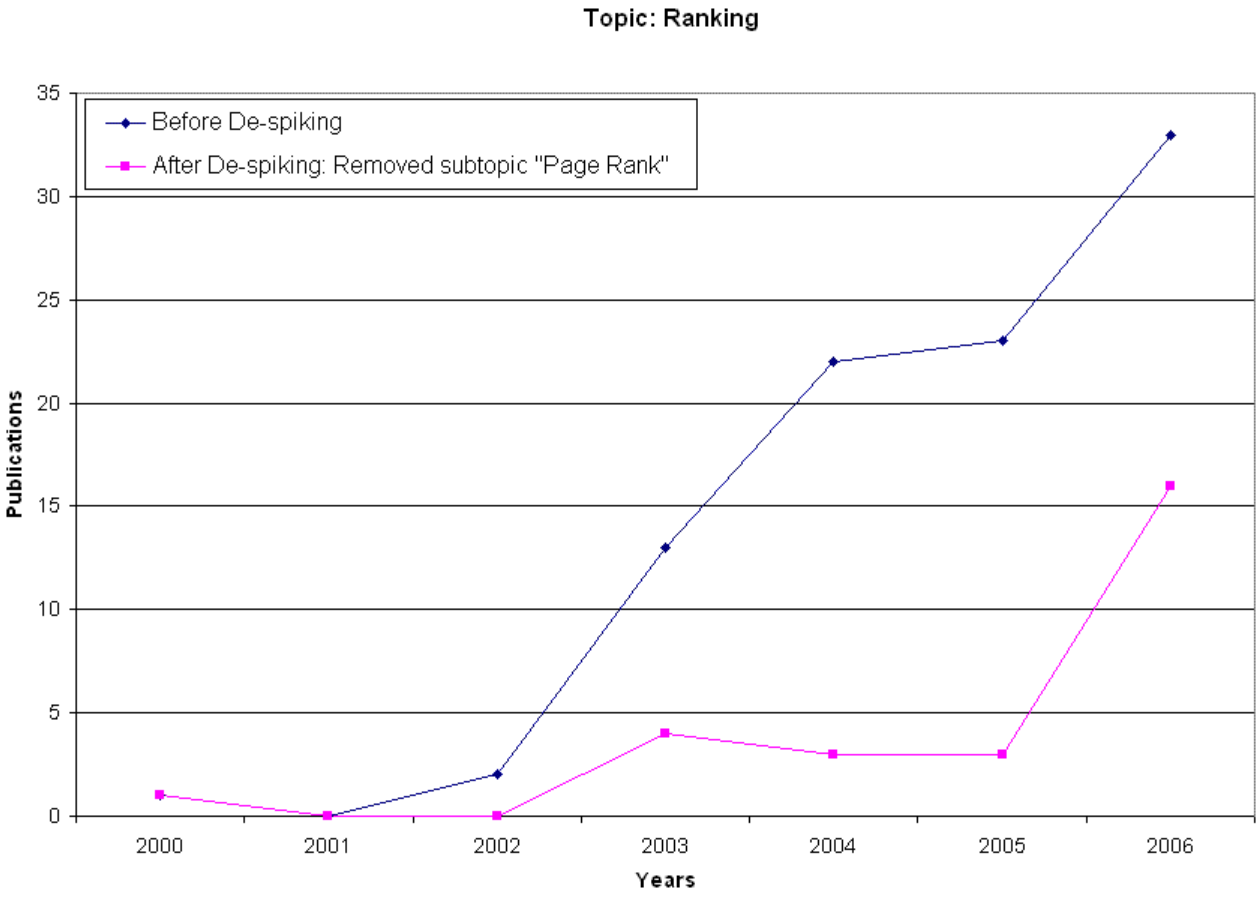


**Figure 9: Example of De-spiking for Bursty Trend Topic: Ranking**

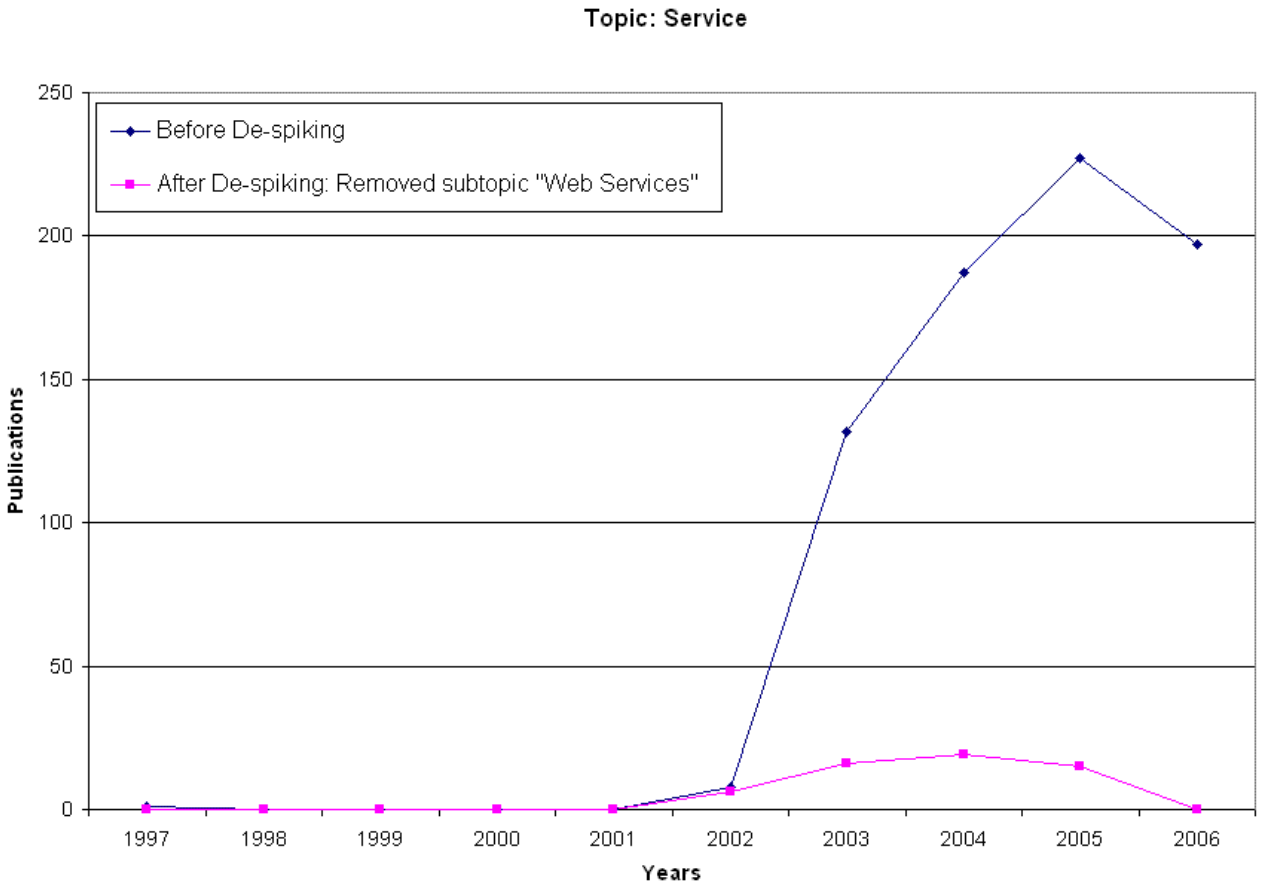


**Figure 10: Example of De-spiking for Bursty Trend Topic: Service**

# CHAPTER 6

# RELATED WORK

The identification of trends has been addressed with techniques such as mining and the use of social networks to name a few. For example, Tho et al. uses a web mining approach for identifying research trends and emergent trends [27]. Their dataset was obtained using Indexing Agents that search for research web sites to download scientific publications. In contrast, our approach uses publications from the DBLP bibliography. Additionally, we use semantics in order to explicitly establish relationships between publications and topics.

Detection of bursts has also been studied in the work of [17]. Subsequent work, detects "bursts" in topic areas based on data extracted from blog feeds through the social network representation by the space of all weblogs [12]. Although our work is similar in the sense that bursts are detected on topics, we deviate from their work when it comes to using a different dataset. Our approach uses metadata of publications. Their approach relies on blog data, which has date/time information at more specific time units than research publications.

In the recent work of Zhou et al. trends of research topics can be found together with indication of how authors impact the topics [29]. However, their main concern is determining how topics are related and where and when these topics evolve. Specifically, they address the question of "Is a newly emergent topic truly new or rather a variation of an old topic"? Our work can complement their work by providing a collection of known "emergent trends" to evaluate.

The creation of taxonomies using web documents has been addressed towards detecting

emerging communities and their associated interests [28]. A difference of our work from such approach is that we do not focus on the building of a taxonomy as the goal. Instead, we aim to demonstrating the value of the paper-to-topics relationships that connect the topics of a taxonomy to research papers.

# CHAPTER 7

# CONCLUSIONS AND FUTURE WORK

In this paper we were able to detect bursty trends and emerging trends using a semantic approach. Both methods used for detection were effective, resulting in detecting 118 research topics as bursty trends and 75 topics as emerging trends from among the listed 344 research topics in our taxonomy. Based on these results, the Computer Science area is indeed evolving in 34% of the topic areas listed in our taxonomy of topics. We were also able to pinpoint several topics that contributed in the burst of specific research topics by means of de-spiking. Our method for identifying researchers in the early stages of a research area was very effective in finding many exact matches of researchers that had major contributions within the research area being identified. We also demonstrated the potential of a semantic approach using only metadata of publications (i.e., abstracts and keywords). It was possible to detect trends without using all content in a document.

Centrality measures were also determined for researchers of publications included in our dataset. However, it was quite clear that there are benefits of using, if available, information of researchers that have more than one name or alias. Without using such ‘same-as’ information of researchers, the computation of centrality values won’t be correct.

For future work, our approach for trend detection could be extended to use the terms of a trend in determining emails that match are related to the terms and the email data could then be used for social network analysis (e.g., identification of communities) possibly relating it back to

authors of papers. In addition, terms of trends of interest could be used for mining or processing other datasets such as intranets, blogs, forums and email corpus. For example, Kleinberg [16] described a scenario of grouping emails by topic of identified trends. Moreover, names in emails could be matched against authors of papers that are related to at trend.

The compilation of metadata from papers based on its keywords and abstracts can be improved. In our work, we found that the information on some publishers' websites was somewhat difficult to extract. Thus, it is possible that the detected *new* terms might not have been new in reality. There are efforts by some publishers to make their information easier to access, such as by means of content feeds in XML. However, they rarely provide all relevant metadata items of a publication. The benefits of making available such information in machine processable formats can lead to better dissemination of the latest publications. Moreover, using richer metadata for determining topics on the field can lead to improved measures of the areas of expertise of researchers. A key aspect in this respect is to assign identifiers (e.g., URIs) for authors of papers towards solving ambiguity issues.

# REFERENCES

[1] Al-Sudani S., Alhulou R., Napoli A., Nauer E.: OntoBib: An Ontology-Based System for the Management of a bibliography, *17th European Conference on Artificial Intelligence*, Riva del Garda, Italy (August 28 -September 3, 2006).

[2] Aleman-Meza, B., Nagarajan, M., Ramakrishnan, C., Ding, L., Kolari, P., Sheth, A.P., Arpinar, I.B., Joshi, A. and Finin, T., Semantic Analytics on Social Networks: Experiences in Addressing the Problem of Conflict of Interest Detection. In *15th International World Wide Web Conference,* (Edinburgh, Scotland, 2006).

[3] Auer, S. and Lehmann, J.: What have Innsbruck and Leipzig in common? Extracting Semantics from Wiki Content. *4th European Semantic Web Conference*. Innsbruck, Austria (2007).

[4] Bizer, C.: D2R MAP—a database to RDF mapping language, *12th International World Wide Web Conference*, Budapest, Hungary (2003).

[5] Brickley, D. and Guha, R.V. Resource Description Framework (RDF) Schema Specification. Proposed Recommendation, World Wide Web Consortium: http://www.w3.org/TR/PR-rdf-schema (1999).

[6] Councill, I. G., Giles, L., Han, H., Manavoglu, E.: Automatic Acknowledgment Indexing: Expanding the Semantics of Contribution in the CiteSeer Digital Library. K-CAP (2005).

[7] Elmacioglu, E., Lee, D.: On Six Degrees of Separation in DBLP-DB and More. *SIGMOD Record*, 34(2):33-40 (June 2005).

[8] Gandon, F. Engineering an ontology for a multi-agent corporate memory system. *Proceedings of ISMICK'01*, 209-228, (2001).

[9] Gevry, D. Detection of emerging trends: Automation of domain expert practices. Master's thesis, Department of Computer Science and Engineering at Lehigh University, 2002

[10] Golbeck, J., Katz, Y., Krech, D., Mannes, A., Wang, T.D, Hendler, J; PaperPuppy: Sniffing the Trail of Semantic Web Publications, *Semantic Web Challenge at ISWC-2006,* Athens, GA, USA (November 2006).

[11] Gruber, T. R. A Translation Approach to Portable Ontology Specifications. *Knowledge Acquisition*, **5**, (2):199-220, 1993.

[12] Gruhl, D., Guha, R., Liben-Nowell, D., Ding, L., Tomkins, A.: Information Diffusion Through Blogspace. *WWW-2004*, New York, New York (May 17-22, 2004).

[13] Halaschek, C., Aleman-Meza, B., Arpinar, I. B., Sheth, A.P. "Discovering and Ranking Semantic Associations over a Large RDF Metabase." *30th International Conference on Very Large Data Bases*, Toronto, Canada, 2004.

[14] Hepp, M. Possible Ontologies: How Reality Constrains the Development of Relevant Ontologies. *IEEE Internet Computing, 11*(1). 90-96, 2007.

[15] Hirsch, J.E. Random samples: Data point – Impact factor. *Science 309*, 1181, 2005.

[16] Kleinberg, J.: Bursty and Hierarchical Structure in Streams. *ISIGKDD '02*, Edmonton, Alberta, Canada (2002).

[17] Kumar, R., Novak, J., Raghavan, P., Tomkins, A.: On the Bursty Evolution of Blogspace. *WWW2003*, Budapest, Hungary, (May 20-24, 2003).

[18] Mika, P. Flink: Semantic Web Technology for the Extraction and Analysis of Social Networks. *Journal of Web Semantics, 3.* 211-223, 2005.

[19] Newman, M.E.J., *Phys. Rev. e Stat. Phys. Plasmas Fluids Relate. Interdiscip. Top*. 64, 016132, 2001.

[20] Nigel Shadbolt, Nicholas Gibbins, Hugh Glaser, Stephen Harris, Monica M. C. Schraefel: CS AKTive Space, or How We Learned to Stop Worrying and Love the Semantic Web. *IEEE Intelligent Systems*, 19(3):41-47 (2004).

[21] Rasmussen, J., Pejtersen, A.M., Schmidt, K. *Taxonomy for cognitive work analysis.* Roskilde, Denmark: Risø National Laboratory. (Risø report M-2871), 1990.

[22] Redmiles, D., Cheng, L., Damian, D., Herbsleb, J., Kellogg, W. Panel: Collaborative Software Engineering – New and Emerging Trends, Supplemental Proceedings Conference on Computer-Supported Work (CSCW 2006 – Banff, Canada), pp. 237-239.

[23] Sheth, A., From Semantic Search & Integration to Analytics. In Dagstuhl Seminar Proceedings 04391, (Dagstuhl, Germany, 2005).

[24] Sheth, A.P., Aleman-Meza, B., Arpinar, I.B., Halaschek, C., Ramakrishnan, C., Bertram, C., Warke, Y., Avant, D., Arpinar, F.S., Anyanwu, K. and Kochut, K. Semantic Association Identification and Knowledge Discovery for National Security Applications. *Journal of Database Management*, *16* (1). 33-53, 2005.

[25] Staab, S., Erdmann, M., Mädche, A., Decker, S., An extensible approach for modeling ontologies in RDF(S), in: First Workshop on the Semantic Web at the Fourth European Conference on Digital Libraries, Lisbon, Portugal, 2000.

[26] The Yahoo! Research Team, “Content, Metadata, and Behavioral Information: Directions for Yahoo! Research”. *IEEE Data Engineering Bulletin*, 31(4): 10-18, (2006).

[27] Tho, Q. T., Hui, S. C., Fong, A.: Web Mining for Identifying Research Trends. *ICADL 2003*, Berlin Heidelberg (2003) 290-301.

[28] Velardi, P., Cucchiarelli, A., Michaël Petit, M.: A Taxonomy Learning Method and its Application to Characterize a Scientific Web Community, *IEEE Transactions on Knowledge and Data Engineering*, 19(2): 180-191 (February 2007).

[29] Zhou, D., Ji, X., Zha, H., Giles, C.L.: Topic Evolution and Social Interactions: How Authors Effect Research. *CIKM-2006*, Arlington, Virginia, USA, pp. 248-257 (2006).

# APPENDIX A

## ONTOLOGY SCHEMA

```
<?xml version="1.0" encoding="UTF-8"?>
<rdf:RDF
  xmlns:owl="http://www.w3.org/2002/07/owl#"
  xmlns:rdf="http://www.w3.org/1999/02/22-rdf-syntax-ns#"
  xmlns:rdfs="http://www.w3.org/2000/01/rdf-schema#"
  xmlns="http://lsdis.cs.uga.edu/projects/semdis/opus#"
  xmlns:dc="http://purl.org/dc/elements/1.1/"
  xmlns:base="http://lsdis.cs.uga.edu/projects/semdis/opus#">


  <owl:Ontology rdf:about="http://lsdis.cs.uga.edu/projects/semdis/opus#">
    <rdfs:label>SwetoDblp Ontology</rdfs:label>
    <rdfs:comment>This file specifies in RDF Schema format the classes and properties for
SwetoDblp.
    These classes and properties are based on the internal LSDIS Library portal engine.
    Contact Person is Boanerges Aleman-Meza (baleman at uga dot edu).
    </rdfs:comment>
    <owl:versionInfo>2006-11-11</owl:versionInfo>
    <dc:creator>Boanerges Aleman-Meza</dc:creator>
  </owl:Ontology>


  <owl:Class rdf:about="http://lsdis.cs.uga.edu/projects/semdis/opus#Article">
    <rdfs:label>Article</rdfs:label>
    <rdfs:subClassOf rdf:resource="http://lsdis.cs.uga.edu/projects/semdis/opus#Publication"/>
    <rdfs:comment>An article from a journal or magazine.</rdfs:comment>
    <owl:equivalentClass
rdf:resource="http://knowledgeweb.semanticweb.org/semanticportal/OWL/Documentation_Onto
logy.owl#Article_in_Journal" />
    <owl:equivalentClass rdf:resource="http://sw-portal.deri.org/ontologies/swportal#Article" />
    <owl:equivalentClass rdf:resource="http://purl.org/net/nknouf/ns/bibtex#Article" />
  </owl:Class>

  <owl:Class rdf:about="http://lsdis.cs.uga.edu/projects/semdis/opus#Article_in_Proceedings">
    <rdfs:label>Article in Proceedings</rdfs:label>
```

```
    <rdfs:comment>An article in the proceedings of a meeting, such as a conference, workhshop
and symposium.</rdfs:comment>
    <rdfs:subClassOf rdf:resource="http://lsdis.cs.uga.edu/projects/semdis/opus#Publication"/>
    <owl:equivalentClass rdf:resource="http://sw-
portal.deri.org/ontologies/swportal#Inproceedings" />
    <owl:equivalentClass rdf:resource="http://purl.org/net/nknouf/ns/bibtex#Inproceedings" />
  </owl:Class>

  <owl:Class rdf:about="http://lsdis.cs.uga.edu/projects/semdis/opus#Book">
    <rdfs:subClassOf rdf:resource="http://lsdis.cs.uga.edu/projects/semdis/opus#Publication"/>
    <rdfs:label>Book</rdfs:label>
    <rdfs:comment>A book with an explicit publisher.</rdfs:comment>
    <owl:equivalentClass rdf:resource="http://www.marcont.org/ontology/marcont.owl#Book" />
    <owl:equivalentClass
rdf:resource="http://knowledgeweb.semanticweb.org/semanticportal/OWL/Documentation_Onto
logy.owl#Book" />
    <owl:equivalentClass rdf:resource="http://swrc.ontoware.org/ontology#Book" />
    <owl:equivalentClass rdf:resource="http://www.aktors.org/ontology/portal#Book" />
    <owl:equivalentClass rdf:resource="http://sw-portal.deri.org/ontologies/swportal#Book" />
    <owl:equivalentClass rdf:resource="http://purl.org/net/nknouf/ns/bibtex#Book" />
  </owl:Class>

  <owl:Class rdf:about="http://lsdis.cs.uga.edu/projects/semdis/opus#Book_Chapter">
    <rdfs:label>Book Chapter</rdfs:label>
    <rdfs:subClassOf rdf:resource="http://lsdis.cs.uga.edu/projects/semdis/opus#Publication"/>
    <rdfs:comment>A part of a book, such as a chapter (or section/preface) and/or a range of
pages.</rdfs:comment>

    <owl:equivalentClass
rdf:resource="http://knowledgeweb.semanticweb.org/semanticportal/OWL/Documentation_Onto
logy.owl#Article_in_Book" />
    <owl:equivalentClass rdf:resource="http://purl.org/net/nknouf/ns/bibtex#Inbook" />
  </owl:Class>

  <owl:Class rdf:about="http://lsdis.cs.uga.edu/projects/semdis/opus#Doctoral_Dissertation">
    <rdfs:comment>A dissertation written to receive a PhD.</rdfs:comment>
    <rdfs:subClassOf rdf:resource="http://lsdis.cs.uga.edu/projects/semdis/opus#Thesis"/>
    <rdfs:label>Doctoral Dissertation</rdfs:label>
    <owl:equivalentClass
rdf:resource="http://knowledgeweb.semanticweb.org/semanticportal/OWL/Documentation_Onto
logy.owl#PhD_Thesis" />
    <owl:equivalentClass rdf:resource="http://swrc.ontoware.org/ontology#PhDThesis" />
    <owl:equivalentClass rdf:resource="http://sw-portal.deri.org/ontologies/swportal#PhDThesis"
/>
```

```
    <owl:equivalentClass rdf:resource="http://purl.org/net/nknouf/ns/bibtex#Phdthesis" />
  </owl:Class>

  <owl:Class rdf:about="http://lsdis.cs.uga.edu/projects/semdis/opus#Edited_Book">
    <rdfs:subClassOf
rdf:resource="http://lsdis.cs.uga.edu/projects/semdis/opus#Edited_Publication"/>
    <rdfs:subClassOf rdf:resource="http://lsdis.cs.uga.edu/projects/semdis/opus#Book"/>
    <rdfs:label>Edited Book</rdfs:label>
    <rdfs:comment>An edited book with an explicit publisher.</rdfs:comment>
    <owl:equivalentClass rdf:resource="http://www.aktors.org/ontology/portal#Edited-Book" />
  </owl:Class>

  <owl:Class rdf:about="http://lsdis.cs.uga.edu/projects/semdis/opus#Edited_Publication">
    <rdfs:label>Edited Publication</rdfs:label>
    <rdfs:subClassOf rdf:resource="http://lsdis.cs.uga.edu/projects/semdis/opus#Publication"/>
    <rdfs:comment>An edited publication, that is, it has one or more editors (edited books,
etc)</rdfs:comment>
  </owl:Class>

  <owl:Class rdf:about="http://lsdis.cs.uga.edu/projects/semdis/opus#Journal">
    <rdfs:subClassOf
rdf:resource="http://lsdis.cs.uga.edu/projects/semdis/opus#Serial_Publication"/>
    <rdfs:comment>A periodical presenting articles on a particular subject.</rdfs:comment>
    <rdfs:label>Journal</rdfs:label>
    <owl:equivalentClass rdf:resource="http://www.aktors.org/ontology/portal#Journal" />
    <owl:equivalentClass rdf:resource="http://sw-portal.deri.org/ontologies/swportal#Journal" />
  </owl:Class>

  <owl:Class rdf:about="http://lsdis.cs.uga.edu/projects/semdis/opus#Masters_Thesis">
    <rdfs:subClassOf rdf:resource="http://lsdis.cs.uga.edu/projects/semdis/opus#Thesis"/>
    <rdfs:comment>A thesis written to receive a Master's degree.</rdfs:comment>
    <rdfs:label>Masters Thesis</rdfs:label>
    <owl:equivalentClass
rdf:resource="http://knowledgeweb.semanticweb.org/semanticportal/OWL/Documentation_Onto
logy.owl#Master_Thesis" />
    <owl:equivalentClass rdf:resource="http://swrc.ontoware.org/ontology#MasterThesis" />
    <owl:equivalentClass rdf:resource="http://sw-
portal.deri.org/ontologies/swportal#MasterThesis" />
    <owl:equivalentClass rdf:resource="http://purl.org/net/nknouf/ns/bibtex#Mastersthesis" />
  </owl:Class>

  <owl:Class rdf:about="http://lsdis.cs.uga.edu/projects/semdis/opus#Proceedings">
    <rdfs:subClassOf rdf:resource="http://lsdis.cs.uga.edu/projects/semdis/opus#Publication"/>
    <rdfs:label>Proceedings</rdfs:label>
```

```
    <rdfs:comment>A written account of what transpired at a meeting.</rdfs:comment>
    <owl:equivalentClass rdf:resource="http://swrc.ontoware.org/ontology#Proceedings" />
    <owl:equivalentClass rdf:resource="http://sw-
portal.deri.org/ontologies/swportal#Proceedings" />
    <owl:equivalentClass rdf:resource="http://purl.org/net/nknouf/ns/bibtex#Proceedings" />
  </owl:Class>

  <owl:Class rdf:about="http://lsdis.cs.uga.edu/projects/semdis/opus#Publication">
    <rdfs:label>Publication</rdfs:label>
    <rdfs:subClassOf rdf:resource="http://xmlns.com/foaf/0.1/Document"/>
    <rdfs:comment>Individual documents and collections of documents such as series, journals,
etc.</rdfs:comment>

    <owl:equivalentClass
rdf:resource="http://knowledgeweb.semanticweb.org/semanticportal/OWL/Documentation_Onto
logy.owl#Publication" />
    <owl:equivalentClass rdf:resource="http://swrc.ontoware.org/ontology#Publication" />
    <owl:equivalentClass rdf:resource="http://www.aktors.org/ontology/portal#Publication" />
    <owl:equivalentClass rdf:resource="http://sw-portal.deri.org/ontologies/swportal#Publication"
/>
  </owl:Class>

  <owl:Class rdf:about="http://lsdis.cs.uga.edu/projects/semdis/opus#Publishing_Organization">
    <rdfs:subClassOf rdf:resource="http://xmlns.com/foaf/0.1/Organization"/>
    <owl:equivalentClass rdf:resource="http://swrc.ontoware.org/ontology#Organization" />
    <rdfs:label>Publisher</rdfs:label>
    <rdfs:comment>An organization that, among other things, creates publishing periodicals,
books or music.</rdfs:comment>
    <owl:equivalentClass rdf:resource="http://www.aktors.org/ontology/portal#Publishing-
House" />
  </owl:Class>

  <owl:Class rdf:about="http://lsdis.cs.uga.edu/projects/semdis/opus#School">
    <rdfs:label>School</rdfs:label>
    <rdfs:subClassOf rdf:resource="http://xmlns.com/foaf/0.1/Organization"/>
    <rdfs:comment>An organization where individuals receive education.</rdfs:comment>
  </owl:Class>

  <owl:Class rdf:about="http://lsdis.cs.uga.edu/projects/semdis/opus#Serial_Publication">

    <rdfs:label>Serial Publication</rdfs:label>
    <rdfs:subClassOf rdf:resource="http://lsdis.cs.uga.edu/projects/semdis/opus#Publication"/>
    <rdfs:comment>A periodical that appears at scheduled times.</rdfs:comment>
  </owl:Class>
```

```
  <owl:Class rdf:about="http://lsdis.cs.uga.edu/projects/semdis/opus#Series">
    <rdfs:comment>Publication Series, such as LNCS, WEUR Workshops, etc. (at this time
debatable whether this should be subclassof Publication)</rdfs:comment>
    <rdfs:label>Series</rdfs:label>

  </owl:Class>

  <owl:Class rdf:about="http://lsdis.cs.uga.edu/projects/semdis/opus#Thesis">
    <rdfs:label>Thesis</rdfs:label>
    <rdfs:subClassOf rdf:resource="http://lsdis.cs.uga.edu/projects/semdis/opus#Publication"/>
    <rdfs:comment>A treatise advancing a new point of view resulting from research; usually a
requirement for an advanced academic degree.</rdfs:comment>
    <owl:equivalentClass
rdf:resource="http://knowledgeweb.semanticweb.org/semanticportal/OWL/Documentation_Onto
logy.owl#Thesis" />
    <owl:equivalentClass rdf:resource="http://swrc.ontoware.org/ontology#Thesis" />
    <owl:equivalentClass rdf:resource="http://sw-portal.deri.org/ontologies/swportal#Thesis" />
  </owl:Class>

  <owl:Class rdf:about="http://lsdis.cs.uga.edu/projects/semdis/opus#University">
    <rdfs:label>University</rdfs:label>
    <rdfs:subClassOf rdf:resource="http://lsdis.cs.uga.edu/projects/semdis/opus#School"/>
    <rdfs:comment>An institution for higher learning with teaching and research facilities
constituting a graduate school and professional schools that award master's degrees and
doctorates and an undergraduate division that awards bachelor's degrees.</rdfs:comment>
    <owl:equivalentClass
rdf:resource="http://knowledgeweb.semanticweb.org/semanticportal/OWL/Organization_Ontolo
gy.owl#University" />
    <owl:equivalentClass rdf:resource="http://swrc.ontoware.org/ontology#University" />
    <owl:equivalentClass rdf:resource="http://www.aktors.org/ontology/portal#University" />
    <owl:equivalentClass rdf:resource="http://sw-portal.deri.org/ontologies/swportal#University"
/>
  </owl:Class>

  <owl:Class rdf:about="http://lsdis.cs.uga.edu/projects/semdis/opus#Webpage">
    <rdfs:comment>A webpage, it is subclass of Document because we want to emphasize that
the URL of the webpage is used the URI.</rdfs:comment>
    <rdfs:subClassOf rdf:resource="http://xmlns.com/foaf/0.1/Document"/>
    <rdfs:label>Webpage</rdfs:label>
  </owl:Class>
  <owl:ObjectProperty rdf:about="http://lsdis.cs.uga.edu/projects/semdis/opus#at_university">
    <rdfs:comment>Indicates that a publication originates or is related to a specific
University.</rdfs:comment>
    <rdfs:label>at university</rdfs:label>
```

```
    <rdfs:range rdf:resource="http://lsdis.cs.uga.edu/projects/semdis/opus#University"/>
    <rdfs:domain rdf:resource="http://lsdis.cs.uga.edu/projects/semdis/opus#Publication"/>
  </owl:ObjectProperty>

  <owl:ObjectProperty rdf:about="http://lsdis.cs.uga.edu/projects/semdis/opus#chapter_of">
    <rdfs:comment>Indicates that a book chapter belongs to a specific book. It is debateable
whether this should be subclass of Collection.</rdfs:comment>
    <rdfs:domain rdf:resource="http://lsdis.cs.uga.edu/projects/semdis/opus#Book_Chapter"/>
    <rdfs:range rdf:resource="http://lsdis.cs.uga.edu/projects/semdis/opus#Edited_Book"/>
    <rdfs:label>Chapter Of</rdfs:label>
  </owl:ObjectProperty>

  <owl:ObjectProperty rdf:about="http://lsdis.cs.uga.edu/projects/semdis/opus#cites">

    <rdfs:comment>Indicates that a publication cites another publication.</rdfs:comment>
    <rdfs:label>Cites</rdfs:label>
    <rdfs:range rdf:resource="http://lsdis.cs.uga.edu/projects/semdis/opus#Publication"/>
    <rdfs:domain rdf:resource="http://lsdis.cs.uga.edu/projects/semdis/opus#Publication"/>
    <owl:equivalentProperty rdf:resource="http://swrc.ontoware.org/ontology#cite" />
  </owl:ObjectProperty>

  <owl:ObjectProperty rdf:about="http://lsdis.cs.uga.edu/projects/semdis/opus#isIncludedIn">
    <rdfs:comment>Indicates that a publication is included in a specific proceedings
publication.</rdfs:comment>
    <rdfs:range rdf:resource="http://lsdis.cs.uga.edu/projects/semdis/opus#Proceedings"/>
    <rdfs:label>is Included in Proceedings</rdfs:label>
    <rdfs:domain
rdf:resource="http://lsdis.cs.uga.edu/projects/semdis/opus#Article_in_Proceedings"/>
    <owl:equivalentProperty rdf:resource="http://sw-
portal.deri.org/ontologies/swportal#containedInProceedings" />
  </owl:ObjectProperty>

  <owl:ObjectProperty rdf:about="http://lsdis.cs.uga.edu/projects/semdis/opus#editor">
    <rdfs:comment>Indicates that a publication has a specific editor(s).</rdfs:comment>

    <rdfs:range rdf:resource="http://www.w3.org/1999/02/22-rdf-syntax-ns#Seq"/>
    <rdfs:domain
rdf:resource="http://lsdis.cs.uga.edu/projects/semdis/opus#Edited_Publication"/>
    <owl:equivalentProperty rdf:resource="http://swrc.ontoware.org/ontology#editor" />
    <rdfs:label>Editor</rdfs:label>
  </owl:ObjectProperty>

  <owl:ObjectProperty rdf:about="http://lsdis.cs.uga.edu/projects/semdis/opus#affiliation">
    <rdfs:comment>Indicates that a person is affiliated to a specific organization.</rdfs:comment>
```

```
    <rdfs:label>Affiliation</rdfs:label>
    <rdfs:domain rdf:resource="http://xmlns.com/foaf/0.1/Person"/>
    <rdfs:range rdf:resource="http://xmlns.com/foaf/0.1/Organization"/>
    <owl:equivalentProperty rdf:resource="http://swrc.ontoware.org/ontology#affiliation" />
    <owl:equivalentProperty rdf:resource="http://www.aktors.org/ontology/portal#has-affiliation"
/>
  </owl:ObjectProperty>

  <owl:ObjectProperty rdf:about="http://lsdis.cs.uga.edu/projects/semdis/opus#in_series">
    <rdfs:comment>Indicates that a Publication is part of a Publication Series.</rdfs:comment>

    <rdfs:label>In Series</rdfs:label>
    <rdfs:range rdf:resource="http://lsdis.cs.uga.edu/projects/semdis/opus#Series"/>
    <rdfs:domain rdf:resource="http://lsdis.cs.uga.edu/projects/semdis/opus#Publication"/>
  </owl:ObjectProperty>

  <owl:ObjectProperty rdf:about="http://lsdis.cs.uga.edu/projects/semdis/opus#author">
    <rdfs:range rdf:resource="http://www.w3.org/1999/02/22-rdf-syntax-ns#Seq"/>
    <rdfs:label>Author</rdfs:label>

    <rdfs:domain rdf:resource="http://lsdis.cs.uga.edu/projects/semdis/opus#Publication"/>
    <rdfs:comment>Indicates that a publication is authored by a specific
person(s).</rdfs:comment>
    <owl:equivalentProperty rdf:resource="http://swrc.ontoware.org/ontology#author" />
  </owl:ObjectProperty>




  <owl:DatatypeProperty rdf:about="http://lsdis.cs.uga.edu/projects/semdis/opus#book_title">
    <rdfs:range rdf:resource="http://www.w3.org/2001/XMLSchema#string"/>

    <rdfs:comment>An alternative Book Title or the Book Title where an article appears, such as
title of proceedings.</rdfs:comment>
    <rdfs:label>book title</rdfs:label>
    <rdfs:domain rdf:resource="http://lsdis.cs.uga.edu/projects/semdis/opus#Publication"/>
    <owl:equivalentProperty rdf:resource="http://swrc.ontoware.org/ontology#booktitle" />
    <owl:equivalentProperty rdf:resource="http://purl.org/net/nknouf/ns/bibtex#hasBooktitle" />
  </owl:DatatypeProperty>

  <owl:DatatypeProperty rdf:about="http://lsdis.cs.uga.edu/projects/semdis/opus#cdrom">
    <rdfs:range rdf:resource="http://www.w3.org/2001/XMLSchema#string"/>
    <rdfs:comment>The CDROM location of a Publication, as used by the ACM SIGMOD
Anthology.</rdfs:comment>
```

```
    <rdfs:label>cdrom</rdfs:label>
    <rdfs:domain rdf:resource="http://lsdis.cs.uga.edu/projects/semdis/opus#Publication"/>
  </owl:DatatypeProperty>

  <owl:DatatypeProperty rdf:about="http://lsdis.cs.uga.edu/projects/semdis/opus#chapter">
    <rdfs:range rdf:resource="http://www.w3.org/2001/XMLSchema#string"/>
    <rdfs:comment>The chapter number of a publication</rdfs:comment>
    <rdfs:label>chapter</rdfs:label>
    <rdfs:domain rdf:resource="http://lsdis.cs.uga.edu/projects/semdis/opus#Publication"/>
    <owl:equivalentProperty
rdf:resource="http://knowledgeweb.semanticweb.org/semanticportal/OWL/Documentation_Onto
logy.owl#Chapter" />
    <owl:equivalentProperty rdf:resource="http://swrc.ontoware.org/ontology#chapter" />
  </owl:DatatypeProperty>

  <owl:DatatypeProperty rdf:about="http://lsdis.cs.uga.edu/projects/semdis/opus#ee">
    <rdfs:range rdf:resource="http://www.w3.org/2001/XMLSchema#string"/>
    <rdfs:comment>The Electronic Edition of a publication</rdfs:comment>

    <rdfs:label>ee</rdfs:label>
    <rdfs:domain rdf:resource="http://lsdis.cs.uga.edu/projects/semdis/opus#Publication"/>
  </owl:DatatypeProperty>

  <owl:DatatypeProperty rdf:about="http://lsdis.cs.uga.edu/projects/semdis/opus#isbn">
    <rdfs:range rdf:resource="http://www.w3.org/2001/XMLSchema#string"/>
    <rdfs:comment>The ISBN of a publication.</rdfs:comment>
    <rdfs:label>isbn</rdfs:label>

    <rdfs:domain rdf:resource="http://lsdis.cs.uga.edu/projects/semdis/opus#Publication"/>
    <owl:equivalentProperty
rdf:resource="http://knowledgeweb.semanticweb.org/semanticportal/OWL/Documentation_Onto
logy.owl#ISBN" />
    <owl:equivalentProperty rdf:resource="http://swrc.ontoware.org/ontology#isbn" />
    <owl:equivalentProperty rdf:resource="http://purl.org/net/nknouf/ns/bibtex#hasISBN" />
  </owl:DatatypeProperty>

  <owl:DatatypeProperty
rdf:about="http://lsdis.cs.uga.edu/projects/semdis/opus#journal_name">
    <rdfs:range rdf:resource="http://www.w3.org/2001/XMLSchema#string"/>
    <rdfs:comment>The name of a Journal, such as where an article appears</rdfs:comment>
    <rdfs:label>journal name</rdfs:label>
    <rdfs:domain rdf:resource="http://lsdis.cs.uga.edu/projects/semdis/opus#Journal"/>
    <owl:equivalentProperty rdf:resource="http://purl.org/net/nknouf/ns/bibtex#hasJournal" />
  </owl:DatatypeProperty>
```

```
  <owl:DatatypeProperty
rdf:about="http://lsdis.cs.uga.edu/projects/semdis/opus#last_modified_date">
    <rdfs:domain rdf:resource="http://xmlns.com/foaf/0.1/Document"/>
    <rdfs:comment>The last modified date of a document.</rdfs:comment>
    <rdfs:range rdf:resource="http://www.w3.org/2001/XMLSchema#string"/>
    <rdfs:label>last modified date</rdfs:label>

  </owl:DatatypeProperty>

  <owl:DatatypeProperty rdf:about="http://lsdis.cs.uga.edu/projects/semdis/opus#month">
    <rdfs:domain rdf:resource="http://xmlns.com/foaf/0.1/Document"/>
    <rdfs:comment>The month part of the date of a foaf:Document.</rdfs:comment>
    <rdfs:range rdf:resource="http://www.w3.org/2001/XMLSchema#string"/>
    <rdfs:label>month</rdfs:label>
    <owl:equivalentProperty rdf:resource="http://swrc.ontoware.org/ontology#month" />
    <owl:equivalentProperty rdf:resource="http://purl.org/net/nknouf/ns/bibtex#hasMonth" />
  </owl:DatatypeProperty>

  <owl:DatatypeProperty rdf:about="http://lsdis.cs.uga.edu/projects/semdis/opus#number">
    <rdfs:range rdf:resource="http://www.w3.org/2001/XMLSchema#string"/>
    <rdfs:comment>The Number part of citation of a publication.</rdfs:comment>
    <rdfs:label>number</rdfs:label>
    <rdfs:domain rdf:resource="http://lsdis.cs.uga.edu/projects/semdis/opus#Publication"/>
    <owl:equivalentProperty rdf:resource="http://purl.org/net/nknouf/ns/bibtex#hasNumber" />
  </owl:DatatypeProperty>

  <owl:DatatypeProperty rdf:about="http://lsdis.cs.uga.edu/projects/semdis/opus#pages">
    <rdfs:range rdf:resource="http://www.w3.org/2001/XMLSchema#string"/>
    <rdfs:comment>The Pages part of citation of a publication.</rdfs:comment>
    <rdfs:label>pages</rdfs:label>
    <rdfs:domain rdf:resource="http://lsdis.cs.uga.edu/projects/semdis/opus#Pages" />
    <owl:equivalentProperty
rdf:resource="http://knowledgeweb.semanticweb.org/semanticportal/OWL/Documentation_Onto
logy.owl#Pages" />
    <owl:equivalentProperty rdf:resource="http://swrc.ontoware.org/ontology#pages" />
    <owl:equivalentProperty rdf:resource="http://www.aktors.org/ontology/portal#has-page-
numbers" />
    <owl:equivalentProperty rdf:resource="http://purl.org/net/nknouf/ns/bibtex#hasPages" />
  </owl:DatatypeProperty>

  <owl:DatatypeProperty rdf:about="http://lsdis.cs.uga.edu/projects/semdis/opus#volume">
    <rdfs:range rdf:resource="http://www.w3.org/2001/XMLSchema#string"/>
    <rdfs:comment>The Volume part of citation of a publication.</rdfs:comment>
    <rdfs:label>volume</rdfs:label>
```

```
    <rdfs:domain rdf:resource="http://lsdis.cs.uga.edu/projects/semdis/opus#Publication"/>
    <owl:equivalentProperty
rdf:resource="http://knowledgeweb.semanticweb.org/semanticportal/OWL/Documentation_Onto
logy.owl#Volume" />
    <owl:equivalentProperty rdf:resource="http://swrc.ontoware.org/ontology#volume" />
  </owl:DatatypeProperty>

  <owl:DatatypeProperty rdf:about="http://lsdis.cs.uga.edu/projects/semdis/opus#year">
    <rdfs:domain rdf:resource="http://xmlns.com/foaf/0.1/Document"/>
    <rdfs:comment>The year part of the date of a foaf:Document.</rdfs:comment>
    <rdfs:range rdf:resource="http://www.w3.org/2001/XMLSchema#string"/>
    <rdfs:label>year</rdfs:label>
    <owl:equivalentProperty rdf:resource="http://swrc.ontoware.org/ontology#year" />
  </owl:DatatypeProperty>

  <owl:DatatypeProperty rdf:about="http://lsdis.cs.uga.edu/projects/semdis/opus#abstract">
    <rdfs:domain rdf:resource="http://xmlns.com/foaf/0.1/Document"/>
    <rdfs:comment>The abstract of a document</rdfs:comment>
    <rdfs:range rdf:resource="http://www.w3.org/2001/XMLSchema#string"/>
    <rdfs:label>abstract</rdfs:label>
  </owl:DatatypeProperty>

</rdf:RDF>
```

# APPENDIX B

## RESEARCH TOPIC FIGURES

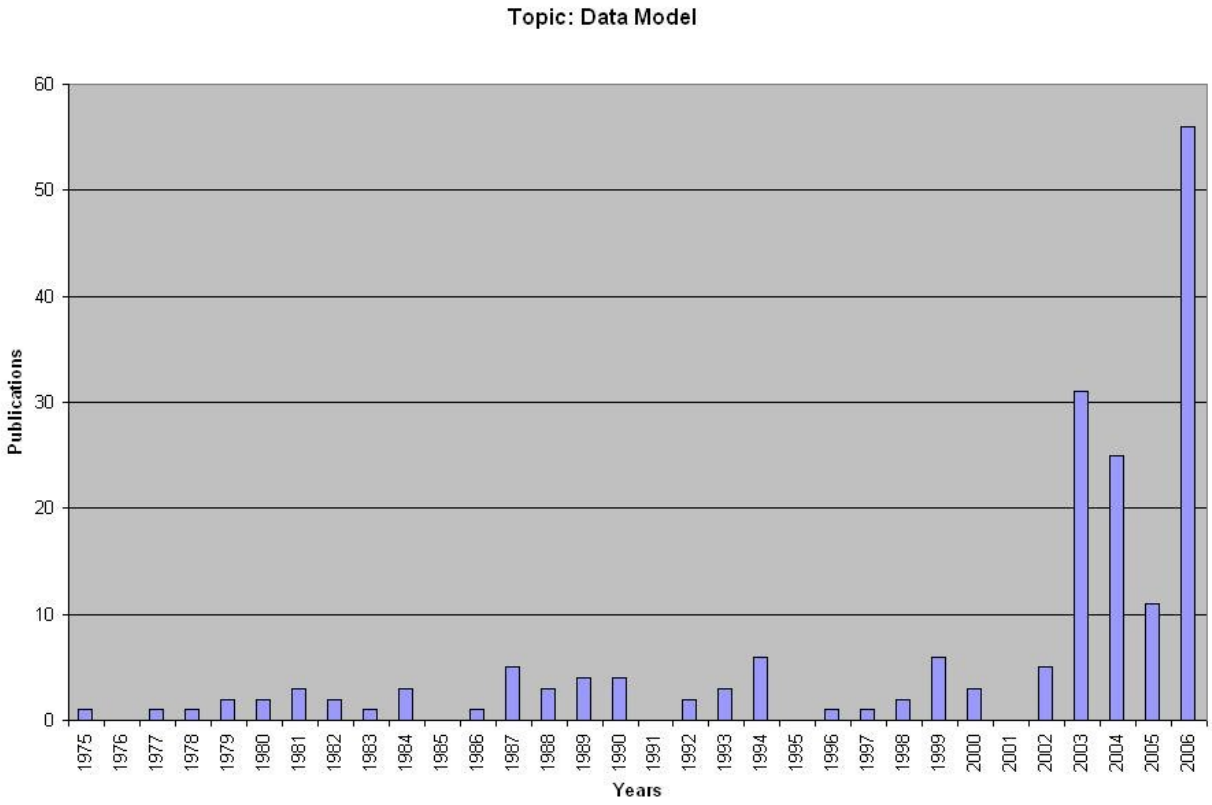


**Figure B.1: Topic: Data Model**

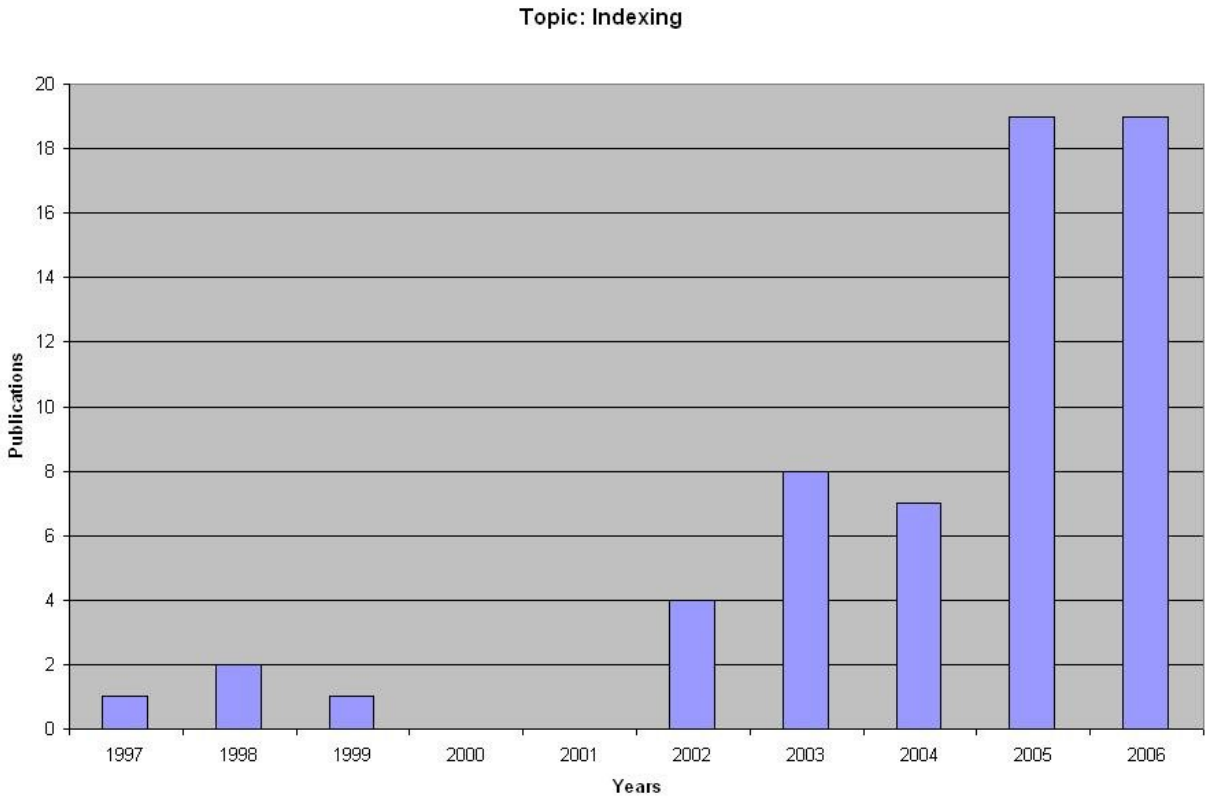


**Figure B.2: Topic: Indexing**

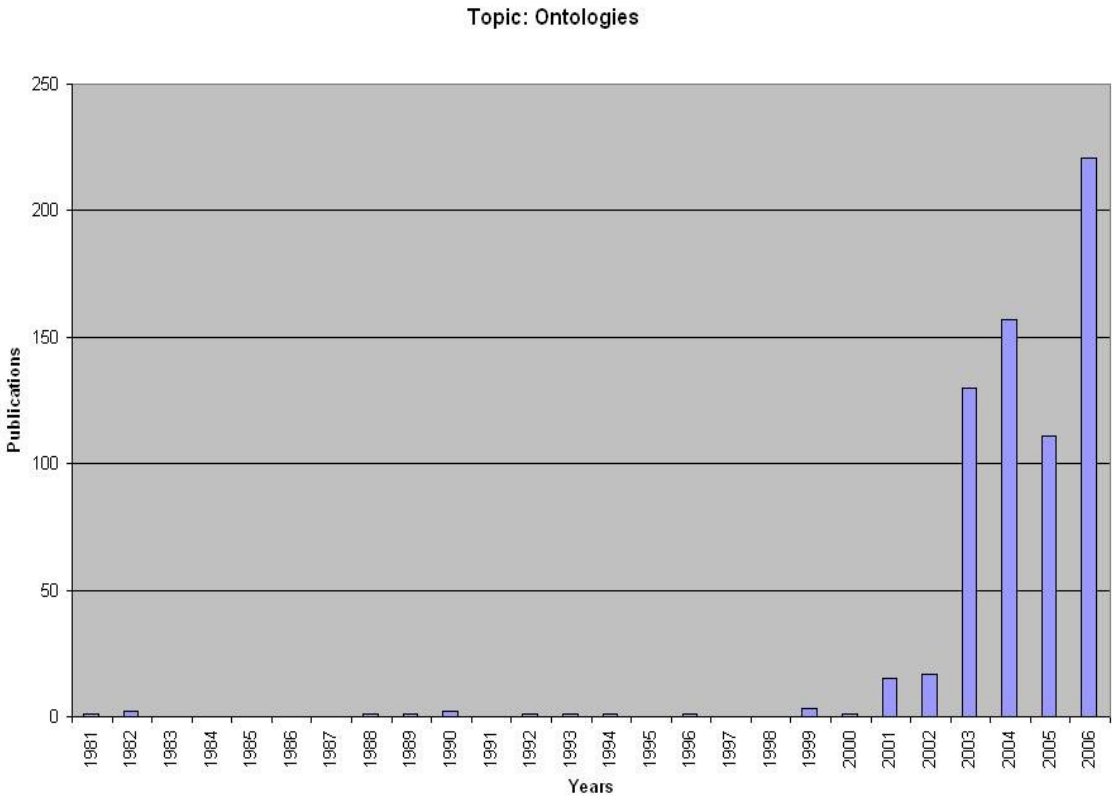


**Figure B.3: Topic: Ontologies**

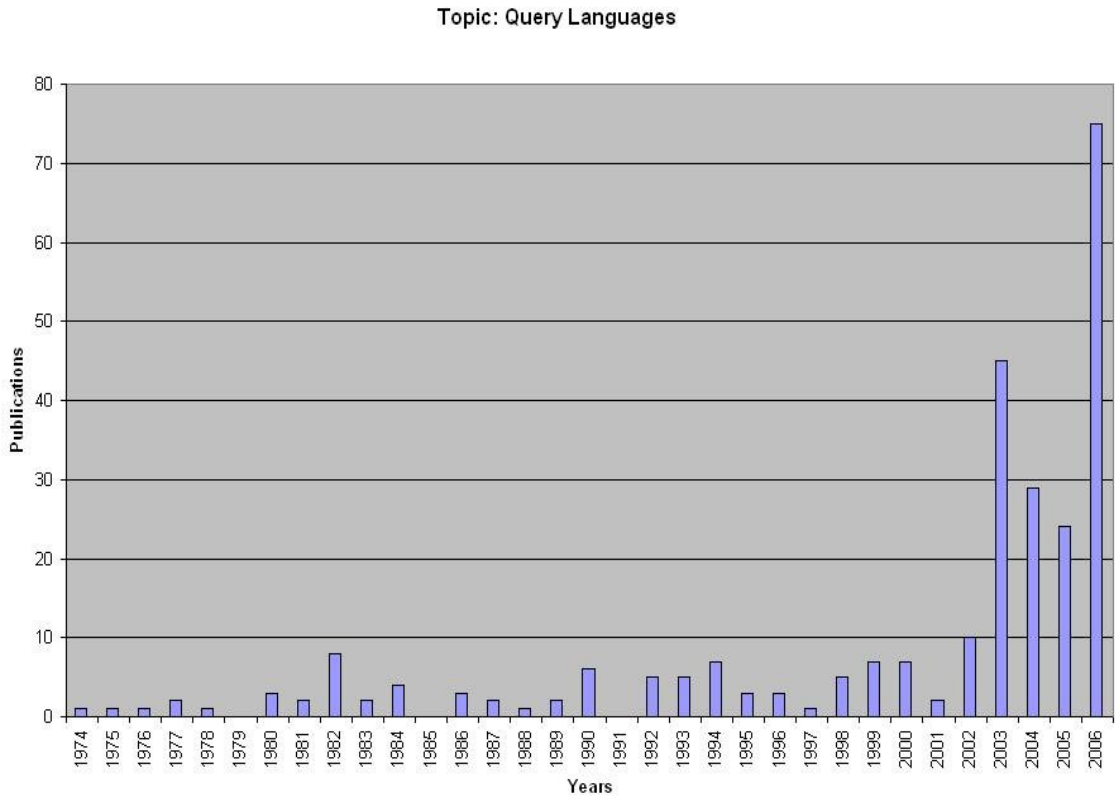


**Figure B.4: Topic: Query Languages**

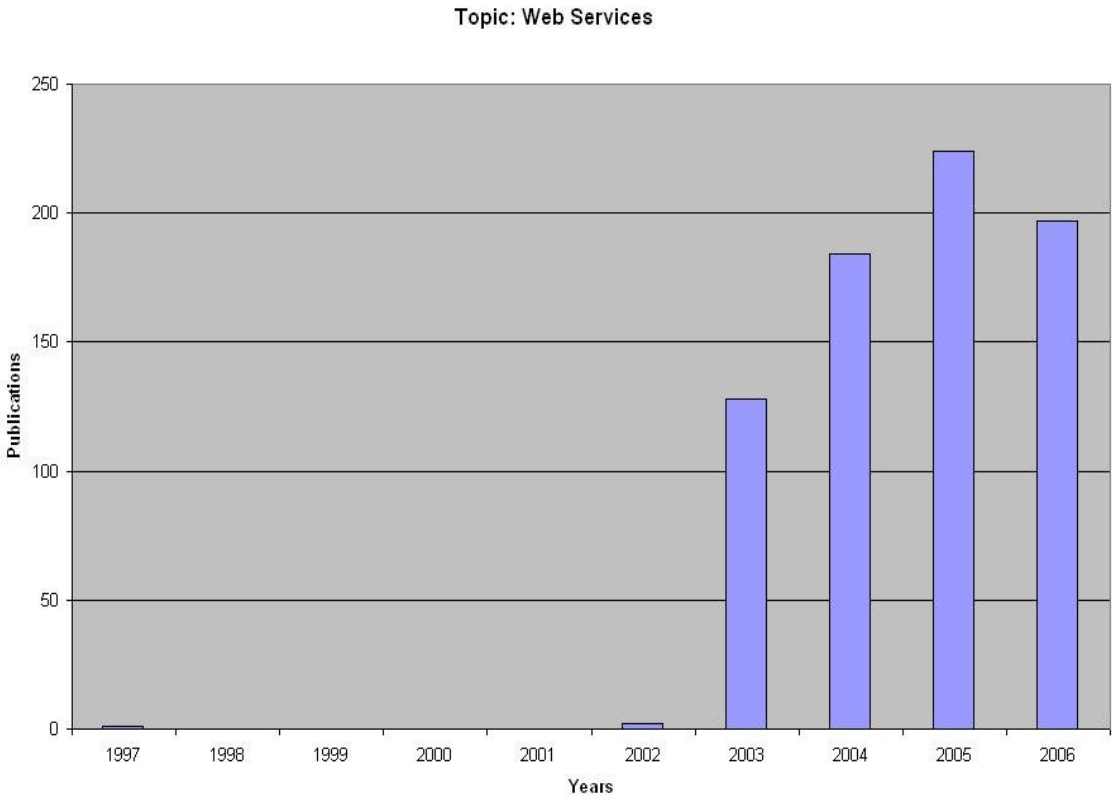


**Figure B.5: Topic: Web Services**

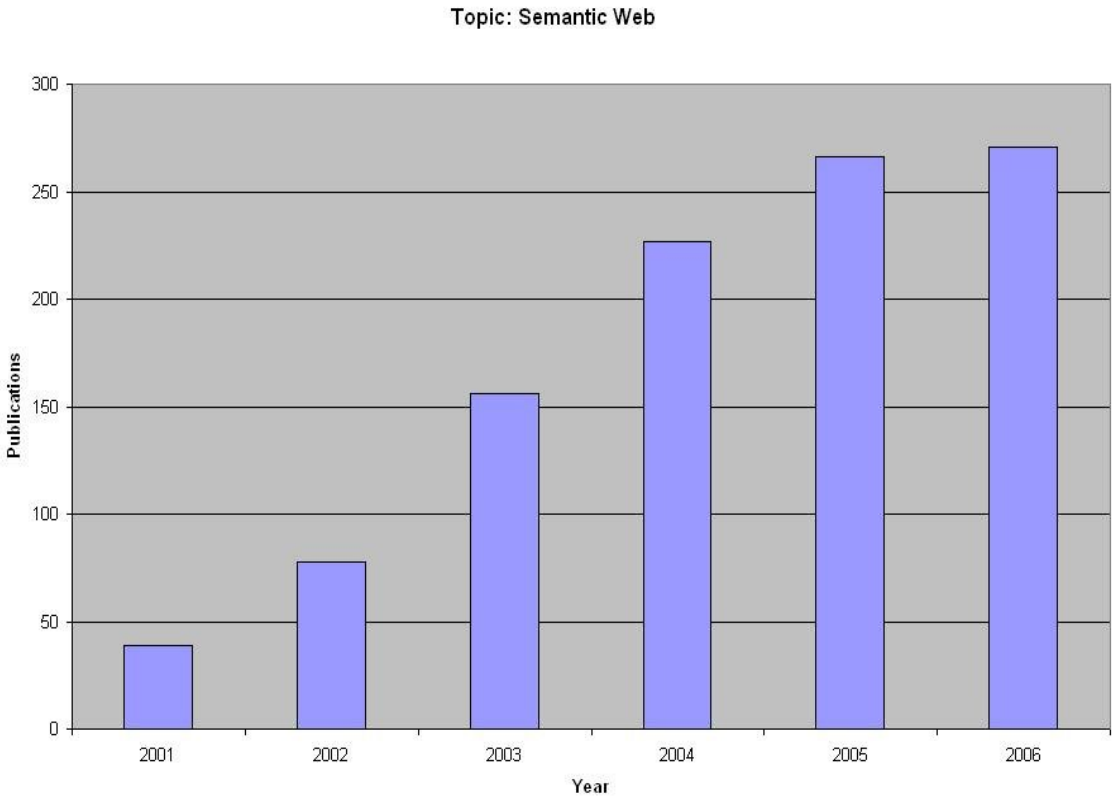


**Figure B.6: Topic: Semantic Web**