International Journal of Engineering Research & Technology (IJERT)
ISSN: 2278-0181
Vol. 2 Issue 3, March - 2013

Text Mining For Expert Witness

Waghode Poonam B
First year Student of M.E (Computer) 2012-13
MIT AOE, ALANDI (D), PUNE

Mayura Kinikar
Assistant Professor of Computer Department
MIT AOE, ALANDI (D), PUNE

***Abstract*—Expert witness is a person who is specialist in a subject, often technical, who may present his/her expert opinion without having been witness to any occurrence relating to the lawsuit. Where text mining is as a process of exploratory data analysis that leads to heretofore unknown information, or to answers for questions for which the answer is not currently known.**

In the publication an application using text mining for attorney is given, if an attorney who need expert advice from an oncologist to determine whether the client was a victim of medical malpractice. Traditionally, I'll have to rely on word-of-mouth recommendations, small directories of self-selected experts on the web or time-consuming research conducted over multiple text collections. Text mining can overcome these problems using Bayesian Belief Network algorithm concept which merge the link entities and records for efficient result to take a decision.

Keywords*— *Text mining, Information Extraction, Record Linkage, Expert Witness

1. I. INTRODUCTION

Suppose an attorney who needed expert advice from an oncologist to determine whether his client was a victim of medical malpractice. How would he find one or more experts, evaluate their ability, and compare them to other experts in the field? Traditionally, he'd have to rely on word-of-mouth recommendations, small directories of self-selected experts on the Web, or time-consuming research conducted over multiple text collections [1].

Due to many coming cases in court and lack of information it's hard for attorney to study on case in good manner. Detail study of any case is the only way to win that case. But this may consume more time, money and lot more things. To help attorneys for their study and information collection, the concept of creating a system that automatically links across documents references to attorneys, judges and expert witnesses [2].

The first step in selecting an expert is to determine exactly what kind of expert an attorney needs. Not all experts are created equal. For example, in the field of psychology, there are about as many different types of experts as there are subspecialties-clinical psychology, neuropsychology, school psychology, child development, child custody, the list could go on forever. The same applies in the field of medicine. So before selecting an expert, first need to consider evidence and try to determine exactly what type of expertise likely will need to achieve legal goals. Do the facts of my case center around a defendant's state of mind at the time of an alleged crime? If so, I'll likely need a mental health expert. Is the case a civil lawsuit where the plaintiff is claiming physical injuries sustained during an accident? If so, we'll likely need a medical expert. Is the only evidence in the case an eyewitness's statement? In this type of situation, I'll likely want to find an expert who can describe the strengths and weaknesses of eyewitness testimony. Attorneys looking for expert witnesses to solidify their positions need to consider several factors when choosing an expert.

So by taking this problem as a main highlight of attorney it's suppose to build expert witness database that links advocate technology to simplify the challenge of finding the best expert witness for a case which will maintain the quality and quantity both from multiple fields.

2. II. LITERATURE SURVEY

Before going the detail of creating database for expert witness, I need some points which are important to understand that are as follows:

A. Trial case summaries

It is a collection of documents that reliably identifies people who have testified as experts in a legal dispute and that provide enough information about them, such documents are shown below. Besides identifying experts who have testified, these documents specify such things as the case's title, the names of participating attorneys and judges, and the factual issues of the case. Figure 1 shows the relevant section of a trial case summary [1].

TITLE: Tenet vs. Frankel
TOPIC: MEDICAL MALPRACTICE
CITY: Cincinnati
STATE: Ohio
COUNTRY: United State
JUDGE: Jennifer Bailey
VERDICT: For the Defendant on January 23, 2002.
ATTORNEYS: …
EXPERTS:
Plaintiff's: Steven Hajdu, M.D., Pathology, Manhasset, NY
Defendant's: Andrew Berchuck, M.D., Gynecological Oncology, Durham, NC
Dominic Maggio, M.D., Internal Medicine, Miami

Figure 1: Part of the trial case summaries

B. Profile record

Profile record contains all the information of expert witness with their details and area of expertise for example for above trial case summaries I preclude profile record as shown in Figure 2 (Profile record).

ANDREW BERCHUCK, MD
GYNECOLOGY/ONCOLOGY
Cincinnati US 27710

AREA(S) OF EXPERTISE:
Medical & Surgical - Obstetrics/Gynecology
Medical & Surgical - Oncology (Cancer)
Medical & Surgical - Physicians/General Practitioners

Figure 2: Profile record

C. Reference record

It is nothing but the records which shows the entities which is to be refer from profile record such as First name, Middle name, Last name, Profession, Organization, City, State as shown below.

First Name: Andrew
Middle Name: Null
Last Name: Berchuck
Profession: Gynecological oncologist
Organization: Null
City: Cincinnati
State: Ohio
Country: US

Figure 3: Reference record for an expert

D. Relationship Summaries

Using the links between documents and entities such as judges, attorneys, and expert witnesses, a new set of products is created that automatically summarize relevant relationships among these entities [2].

E. Cross Document Co-reference Resolution

By combining information extraction techniques with record linkage techniques, I have been able to resolve cross document references for attorneys, judges, and expert witnesses. My basic technique involves extracting a template record for an individual from a text document and matching the template record to an authority file record [2]. Figure 4 depicts how extraction and record linkage are combined to create cross document links.

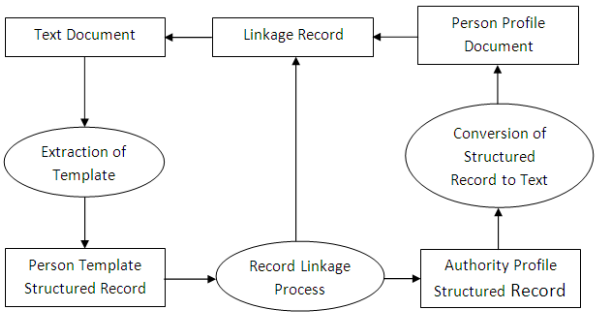

Figure 4: Cross Document Co reference Resolution Process

3. IV. Text Mining

Text mining is as a process of exploratory data analysis that leads to heretofore unknown information, or to answers for questions for which the answer is not currently known [3]. The general idea of text mining – getting small "nuggets" of desired information out of "mountains" of textual data without having to read it all . Text mining is different from what we're familiar with in web search. In search, the user is typically looking for something that is already known and has been written by someone else. The problem is pushing aside all the material that currently isn't relevant to your needs in order to find the relevant information. In text mining, the goal is to discover heretofore unknown information, something that no one yet knows and so could not have yet written down [4].

4. V. Methodology

To create the expert-witness database, I first use information extraction techniques to identify entities and relationships in text. I then use record linkage techniques to attach the entities to database records that stand for real-world people, places, and organizations. Figure 5 shows the application's overall architecture.

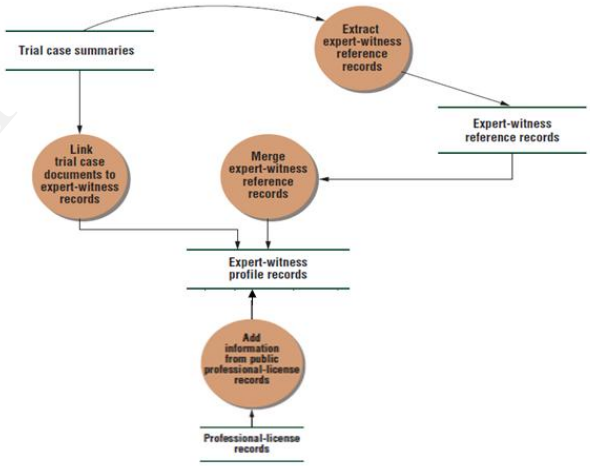

Figure 5: The expert witness application's architecture

A. Information Extraction

Information Extraction is the mapping of natural language texts (such as newswire reports, newspaper and journal articles, electronic mail, World Wide Web pages, any textual database, etc.) into predefined, structured representation, or templates, which, when filled, represent an extract of key information from the original text [7]. The information concerns entities of interest in the application domain (e.g. companies or persons), or relations between such entities, usually in the form of events in which the entities take part (e.g. company takeovers, management successions etc.). Once extracted, the information can then be stored in databases to be queried.

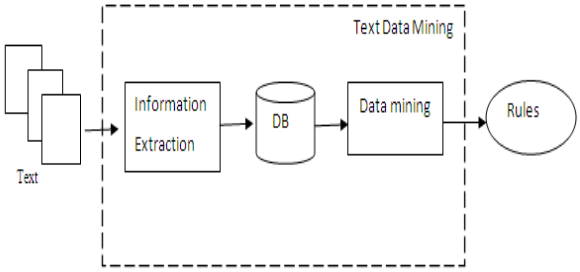

Figure 6: Overview of IE-based text mining framework

The information extraction part of my application has four subcomponents:

- Identification of relevant text in a document,
- Sentence-level information extraction,
- Within-document coreference resolution, and
- Structured event extraction across sentences.

B. Record Linkage

The record linkage portion of the system uses a Bayesian network to match and link attorney, judge, and expert templates to records. This network computes the probability that a given biographical record matches the same person specified in an extracted template. To compute this match likelihood, record linkage is the task of quickly and accurately identifying records corresponding to the same entity from one or more data sources. Record linkage has applications in customer systems for marketing, customer relationship management, fraud detection, data warehousing, law enforcement and government administration. These applications can be classed as 'administrative', because the record linkage is used to make decisions and take actions regarding an individual entity. The central problem that the linking program addresses is determining .whether two names in different reference records from different documents refer to the same person, given the rendition of the names and any contextual information. For example, does the expert-witness reference record of oncologist James Jackson of Palm Springs, California really refer to the same person in the reference record for oncologist James P. Jackson of Sacramento, California? To link mined entities, I use the entity's name and additional evidence from the document.

5. VI. Applications

Text mining has applications in all parts of the research process from literature review and hypothesizing, through experimentation and analysis to generalization, peer review and publishing application categories include:

- Enterprise Business Intelligence/Data Mining, Competitive Intelligence
- E-Discovery, Records Management
- National Security/Intelligence
- Scientific discovery, especially Life Sciences
- Sentiment Analysis Tools, Listening Platforms
- Natural Language/Semantic Toolkit or Service
- Publishing
- Automated ad placement
- Search/Information Access
- Social media monitoring

6. VII. Advantages & Disadvantages

A. Advantages

- It can find information within short period of time with good efficiency and convenience which is far more accessible.
- I get information out of unstructured data stored in some database

B. Drawback/Limitations

- Exact Accuracy of result is missing.
- There is more scope to further extend directory capability, using other application of this technology.
- Further research is needed to establish its effectiveness.

7. VIII. Conclusion

Expert-Witness directory show that even the current state of the art can create value in the form of new databases that gather and disseminate otherwise hard to find information. In the case of experts, the profiles are created by mining references from highly structured and trustworthy documents such as jury verdict and settlement documents and professional license records.

By combining information extraction and record linkage techniques, a creation of a repository of references to attorneys, judges, and expert witnesses across a broad range of text sources. These text sources include caselaw, law reviews, and legal briefs among others.

8. IX. Future Scope

The expert witness database is created using text mining using the concept of merging entities by applying Bayesian belief network algorithm for creating Expert-Witness Database which makes data about expert witness far more accessible to customers. Also described the precision and recall of the programs that create cross document co-references for attorneys and judges in [1] and that create co-references for experts in [2]. Nevertheless, successful applications such as my expert-witness directory show that even the current state of the art can create value in the form of new databases that gather and disseminate otherwise hard-to-find information. I expect to further extend my directory capability, using other applications of this technology. More broadly, text mining could automate many aspects of document processing, such as conceptual indexing, consistency checking, and other laborious manual tasks, although further research is needed to establish its efficacy.

ACKNOWLEDGEMENT

I would like to express my gratitude to all those who have helped me to complete this work. I want to thank my guide, Mrs. Mayura Kinikar, for her continuous help and generous assistance.

She helped in a broad range of issues from giving me direction, helping to find the solutions, outlining the requirements and always having the time to see me.

I have furthermore to thank. Prof. R. M. Goudar, M.E. Coordinator Department of Computer Engineering, to encourage me to go ahead and for continuous guidance. I also want to thank Prof. Uma Nagraj for all her assistance and valuable guidance.

9. REFERENCES

[1] Christopher Dozier, Peter Jackson, "*Mining Text for Expert Witnesses*," IEEE Software, vol. 22, no. 3, pp. 94-100, May-June 2005, doi:10.1109/MS.2005.70

[2] Dozier, C. and Haschart, R., (2000) *Automatic Extraction and Linking of Person Names in Legal Text*. In Proceedings of RIAO-2000 pages 1305-1321, Paris, France.

[3] Hearst, M.A. (1999) "*Untangling text mining*." Proc Annual Meeting of the Association for Computational Linguistics ACL99. University of Maryland, June.

[4] A.H.Tan. *Text mining: The state of the art and the challenges*. Proceedings of the PAKDD 1999 Workshop on Knowledge Discovery from Advanced Databases, pages 65–70, 1999.

[5] Appelt, D. E., Hobbs, J. E., Bear, J., Israel, D., Tyson, M. (1993). FASTUS: *A Finite-State Processor for Information Extraction from Real- World Text*, pages 1172-1178.

[6] Fellegi, I. P., and A. B. Sunter, "*A Theory for Record Linkage*" Journal of the American Statistical Association, 64, 1183-1210.

[7] Winkler, W. E. (2002), "*Record Linkage and Bayesian Networks*" American Statistical Association, Proceedings of the Section on Survey Research Methods, (also at http://www.census.gov/srd/www/byyear.html).

[8] Copas, J. R., and F. J. Hilton (1990), "*Record Linkage: Statistical Models for Matching Computer Records*" Journal of the Royal Statistical Society, A, 153, pp. 287-320.

[9] Sebastiani, F., Text categorization. In Zanasi, A., editor, *Text Mining and its Applications*, pages 109–129, Southampton, UK: WIT Press, 2005.

[10] University of Manchester, "*University enters collaboration to develop text mining applications*" 2011: http://www.manchester.ac.uk/aboutus/news/display/?id=7627.