

Chapter 20

Information Extraction for Call for Paper

Laurent Issertial

Osaka Prefecture University, Japan

Hiroshi Tsuji

Osaka Prefecture University, Japan

ABSTRACT

This paper proposes a system called CFP Manager specialized on IT field and designed to ease the process of searching conference suitable to one's need. At present, the handling of CFP faces two problems: for emails, the huge quantity of CFP received can be easily skimmed through. For websites, the reviewing of some of the main CFP aggregators available online points out the lack of usable criteria. This system proposes to answer to these problems via its architecture consisting of three components: firstly an Information Extraction module extracting relevant information (as date, location, etc...) from CFP using rule based text mining algorithm. The second component enriches the now extracted data with external one from ontology models. Finally the last one displays the said data and allows the end user to perform complex queries on the CFP dataset and thus allow him to only access to CFP suitable for him. In order to validate the authors' proposal, they eventually process the well-known precision / recall metric on our information extraction component with an average of 0.95 for precision and 0.91 for recall on three different 100 CFP dataset. This paper finally discusses the validity of our approach by confronting our system for different queries with two systems already available online (WikiCFP and IEEE Conference Search) and basic text searching approach standing for searching in an email box. On a 100 CFP dataset with the wide variety of usable data and the possibility to perform complex queries we surpass basic text searching method and WikiCFP by not returning the false positive usually returned by them and find a result close to the IEEE system.

DOI: 10.4018/978-1-7998-0951-7.ch020

INTRODUCTION

With today's technology any type of information accessibility has greatly increased and this is also true for our domain of study: Call For Papers (CFP). It has never been so easy to find information on conference usually via numerous websites or simply by email. However, this profusion of information has also its downside. Handling CFP can be seen as a really time consuming task. Regarding the emails, many CFP are received every day and most of them will only be skimmed through or forgotten by lack of time or attention. As for the web, it is possible to find many websites where you can browse a wide range of CFPs such as WikiCFP (2014) or ConferencePartner (2014), or others bound to different institutions like IEEE (2014) or ACM (2012). Nevertheless these sites have several weak points. They do not allow the user to use CFPs obtained by external sources (e.g. Email) ; thus a new research have to be done on a totally different database, sometimes with less information you could actually find in your CFP email. Moreover, conferences research parameters can be really basic and not handle a request more complex than searching different strings of characters (e.g. conference about X topic in Y country with a Z deadline). Those are the issues we propose to handle with our system: CFP Manager.

In one of our previous work (Issertial & Tsuji, 2011), we proposed the concept of a text mining system able to extract relevant information from a group of CFP. In another one concerning visualized comparison of CFP datasets (Issertial, Saga, & Tsuji, 2012), we introduced the idea of extracted data enrichment via the utilization of ontology, another one (Issertial & Tsuji, 2013) focused on the query and interface system along with the different kind of output that can be obtained from the set of relevant data previously collected. This paper proposes to focus on the case of a final user who's looking for information about conferences suited to him and how our proposed system will ease this process. We will do so by reviewing current web-services proposing this functionality and comparing their theoretical results with our proposed system on different queries. This will be introduced along with the description of our system and numerical experimentations proving its good working. CFP will be subject to rule based text mining algorithms in order to extract relevant information. These very data will be enriched via ontology models with concepts close to the ones found in the CFP. Finally, using the previously collected data, we form the base of a user interface system with an intuitive query system allowing the user to perform complex queries.

PROBLEM DESCRIPTION CFP HANDLING PROBLEM

Our problematic can be divided in three parts; how to extract, manage and display relevant data. CFP is already an active subject of research to make the handling of CFP easier. Nevertheless these researches focus on displaying (Sacco, 2005), gather (Jaimez-Gonzalez & Luna-Ramirez, 2013) and creating end-user application (Xin, Li, & Tang, 2007) or web crawler (Yang, 2009) from available sources and do not take an Information Extraction approach as we intend to do. The Information extraction field is a well-known area and many techniques and algorithms have been used to solve this kind of problem like classifiers (WHISK), sequence model (Hidden Markov Chain) or written rules (NLP algorithm) (Feldman & Sanger, 2007) historically support by completion based conferences as MUC (Message Understanding Conference) (Chinchor & Sundheim, 1993). The application of IE to CFP can be traced back up to 1998 (Lazarinis, 1998). From this time researches focusing on extracting relevant information from CFP applying different algorithms as SVM (Support Vector Machine) (Ireson et al., 2005), CRF

14 more pages are available in the full version of this document, which may be purchased using the "Add to Cart" button on the publisher's webpage:
www.igi-global.com/chapter/information-extraction-for-call-for-paper/239946

Related Content

Dependency Parsing in Bangla

Utpal Garain and Sankar De (2014). *Computational Linguistics: Concepts, Methodologies, Tools, and Applications* (pp. 1536-1549).

www.irma-international.org/chapter/dependency-parsing-in-bangla/108792

Understanding and Reasoning with Text

M. Anne Britt, Katja Wiemer, Keith K. Millis, Joseph P. Magliano, Patty Wallace and Peter Hastings (2012). *Cross-Disciplinary Advances in Applied Natural Language Processing: Issues and Approaches* (pp. 133-154).

www.irma-international.org/chapter/understanding-reasoning-text/64585

The Internet of Things and Beyond: Rise of the Non-Human Actors

Arthur Tatnall and Bill Davey (2020). *Natural Language Processing: Concepts, Methodologies, Tools, and Applications* (pp. 1286-1297).

www.irma-international.org/chapter/the-internet-of-things-and-beyond-rise-of-the-non-human-actors/239991

Evaluation of Narrative and Expository Text Summaries Using Latent Semantic Analysis

René Venegas (2012). *Applied Natural Language Processing: Identification, Investigation and Resolution* (pp. 531-544).

www.irma-international.org/chapter/evaluation-narrative-expository-text-summaries/61069

A Tale of Transitions: The Challenges of Integrating Speech Synthesis in Aided Communication

Martine Smith, Janice Murray, Tetzchner Stephen von and Pearl Langan (2010). *Computer Synthesized Speech Technologies: Tools for Aiding Impairment* (pp. 234-256).

www.irma-international.org/chapter/tale-transitions-challenges-integrating-speech/40869