

{tag}

Recent Trends in Mobile and Cloud Computing

{/tag}

IJCA Proceedings on National Conference on

© 2015 by IJCA Journal

NCRMC 2015 - Number 2

Year of Publication: 2015

Authors:

Akhil M. Jaiswal

Anjali B. Raut

{bibtex}ncrmc2919.bib{/bibtex}

## Abstract

A person may have multiple personal name aliases and that same thing might available be on the web. Identifying aliases of a name is useful in information retrieval, investigating about things, knowledge management, sentiment analysis, relation extraction and name disambiguation. Extracting aliases of an entity is important task in various jobs such as identification of relations among entities, web search and entity disambiguation. The objective of detecting aliases from the web is to retrieve all the information pertaining to a personal name whose content is described with different nick names called aliases in different documents of web. The Web contains aliases of popular personalities in various domains like sports, politics,

music, cinema etc. , and the problem is that it does not contain alias information or information related to its alias names. Recently, there are proven methods of extracting aliases through lexical pattern based retrieval tested using real-world name-alias pairs in Japanese and English as training data related to limited domains. In this paper, a method is proposed in which a lexical-pattern-based approach to extract aliases of a given name that is helpful in information retrieval. A set of names is used and their aliases to extract lexical patterns that describe numerous ways in which information related to aliases of a name is presented on the web the aliases which will be extracted using the proposed method can be successfully utilized in an information retrieval task and will improve in a relation detection task.

### Refer

### ences

- . R. Guha and A. Garg, "Disambiguating People in Search,"technical report, Stanford Univ. , 2004.
- . J. Artilles, J. Gonzalo, and F. Verdejo, "A Testbed for People Searching Strategies in the WWW," Proc. SIGIR &apos;05, pp. 569-570, 2005.
- . P. Cimano, S. Handschuh, and S. Staab, "Towards the Self-Annotating Web," Proc. Int&apos;l World Wide Web Conf. (WWW &apos;04),2004.
- . A. Baggaand B. Baldwin, "Entity-Based Cross-Document Coreferencing Using the Vector Space Model," Proc. Int&apos;l Conf. Computational Linguistics (COLING &apos;98), pp. 79-85, 1998. (C : 240)
- . G. Mann and D. Yarowsky, "Unsupervised Personal Name Disambiguation," Proc. Conf. Computational Natural Language Learning (CoNLL&apos;03), pp. 33-40, 2003. (C : 206)
- . R. Bekkerman and A. McCallum, "Disambiguating Web Appearances of People in a Social Network," Proc. Int&apos;l World Wide Web Conf. (WWW &apos;05), pp. 463-470, 2005. (C : 166)
- . C. Galvez and F. Moya-Anegon, "Approximate Personal Name Matching through Finite-State Graphs," J. Am. Soc. for Information Science and Technology, vol. 58, pp. 1-17, 2007. (C : 26)
- . M. Bilenko and R. Mooney, "Adaptive Duplicate Detection Using Learnable String Similarity Measures," Proc. SIGKDD &apos;03, 2003. (C : 418)
- . G. Salton and M. McGill, Introduction to Modern,Information Retrieval. McGraw-Hill Inc. , 1986.
- . M. Mitra, A. Singhal and C. Buckley, "Improving Automatic Query Expansion," Proc. SIGIR &apos;98, pp. 206-214, 1998.

### Index Terms

Computer Science

Algorithms

**Keywords**

Information Retrieval Web Mining Information Extraction Web Text Analysis.