

Query by Analogical Example

Relational Search Using Web Search Engine Indices

Lohit Jain¹

Advisor: Dr. Amitabha Mukerjee¹
{lohit,amit}@iitk.ac.in

¹ Dept. Of Computer Science and Engineering

October 22, 2013

1. Introduction

Relational Search is an effective way to obtain information in an unknown field for users. For example, if an *Apple* user searches for *Microsoft* products, similar *Apple* products are important clues for the search.

As most proposed relational search requires a huge text corpus to be downloaded from the Web, the results are not upto date and the corpus has high construction cost. I plan to implement methods for relational search by using Web search indices which ensures efficiency in terms of construction cost and relevance.

2. The Problem

The problem is to find results based on relation between two named entities and discovering another named entity having similar relation to a third given entity.

For example, if the a person inputs *chocolate* and *cacao* as the first relation and gives *Wine* as the third entity, the system should give *grape* as result.

Another example, *Youtube* is owned by *Google* and *Oracle* by *Sun Microsystems*. If we input *Youtube*, *Google* and *Oracle* we should get output *Sun Microsystems*.

3. Proposed Solution

My proposed solution involves two steps:

Step 1: Finding Relation Extractor.

Step 2: Extracting and ranking terms based on Relational Similarity.

I propose to implement the relation extractor based on *Term Co-occurrence* and if possible on *Lexico-Syntactic Patterns*.

The term co-occurrence method finds terms in search queries which represent the relation between the search querie's entities. For example for relation between *Australia* and *Parliament* may get term for its location, *Canberra*.

The ranking will be based on method used by Kato and Ohshima in [1]. The ranking will generate ranked outputs for each input.

Formalising the problem:

Let (a, b, c) be the given named entities, the proposed system will involve querying the Web for extracting relation from a, b and finding a d which satisfies the same relation with c .

The step 1 is finding relation between a and b .

The step 2 is generating d 's having similar relation to c and ranking them according to relevance.

4. Evaluation

The evaluation will be based on the test corpus of Kato and Ohshima in [1]

For each pairs in test corpus, the input will be a, b, c and expected output d . The topmost suggestions which matches with the expected output will be an evaluation technique.

5. Ethics and Resourcing

The proposed solution requires Web as a resource at computation stage and the test set used by Kato and Ohshima in [1] at the evaluation stage.

There are no additional data sets or corpus required.

References

- [1] Kato, Makoto P., et al. "Query by analogical example: relational search using web search engine indices." Proceedings of the 18th ACM conference on Information and knowledge management. ACM, 2009.
- [2] Yao, Limin, et al. "Structured relation discovery using generative models." Proceedings of the Conference on Empirical Methods in Natural Language Processing. Association for Computational Linguistics, 2011.