Relation Completion using a Perceptive Search Procedure

V. Latha and Dr.T. Hemalatha

Abstract--- The standard way of reconstructing the entities will involve joining the tables. Because of the autonomous and decentralized way in which the sources are populated, they often do not have Primary Key - Foreign Key relations. To answer queries effectively, we need to integrate the information about the individual entities that are fragmented over multiple sources.

For instance, a pattern-based method (PaRE) uses extracted patterns as the auxiliary information in formulating search queries. However, high-quality patterns may decrease the probability of finding suitable target entities. As an alternative, a CoRE method is proposed that uses context terms learned surrounding the expression of a relation as the auxiliary information in formulating queries. The experimental results based on several real-world web data collections demonstrate that CoRE reaches a much higher accuracy than PaRE for the purpose of RC.

Keywords--- Context-Aware Relation Extraction, Relation Completion, Pattern Based Method

I. INTRODUCTION

Relation Completion (RC) is focused on satisfying precise, narrow, pre-specified requests.

The RC Problem is related to some well-studied problems in the areas of data management and information extraction.

The RC problem is studied under the many-to-many mapping and techniques are investigated for maintaining the high precision and recall achieved under the many-to-one case.

In existing system only semantic relations are identified. It does not consider consolidating and pruning those individual candidate sets into a minimal global set of RelQueries. However they only consider one- to- many mapping. Most of the existing schemes are insufficient for Big Data applications because of data placements in varied computation nodes.

Adding link analysis and multiple text representations of documents to existing document ranking functions meant that the internal algorithm of an IR system was becoming complex.

II. BACKGROUND AND CORE OVERVIEW

Relation completion is rapidly becoming one of the fundamental tasks underlying many of the emerging applications that capitalize on the opportunities provided by the abundance of big data (e.g., entity reconstruction [1],[2],data enrichment [3],[4],etc.). We formally define the relation completion task as follows.

Definition 1 (Relation completion)

Given two entity lists Lα and Lβ and a semantic relation R, the goal of relation completion is to identify for each entity α€Lα an entity β€ Lβ which satisfies (α, β) € R. Accordingly, Lα is a query list, Lβ is a target list, α is a query entity and β is α's target entity(Fig 1).

V. Latha, Research Scholar, School of Computing Sciences, Vels University
Dr.T. Hemalatha, Associate.Prof, Dept of M.C.A, School of Computing Science, Vels University.

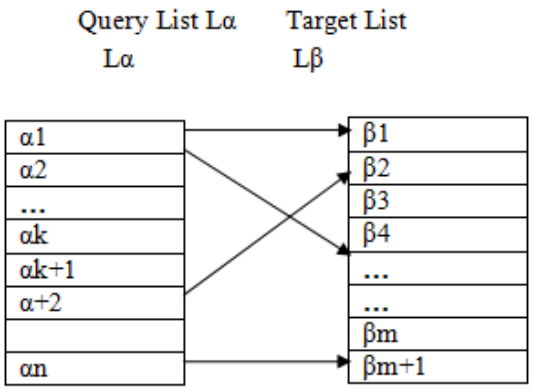


Fig. 1

Definition 2 (Relation Query)

A Relation Query is a web search query formulated to retrieve documents containing the target entity β for the query entity α using some auxiliary information Aux.

Further, a retrieved document that contains the correct target entity is denoted as RelDoc, which is defined as follows:

Definition 3 (Relation-Cotext Document (RelDoc))

A retrieved document is denoted Relation-Context Document if and only if it contains the target entity β for the query entity α.

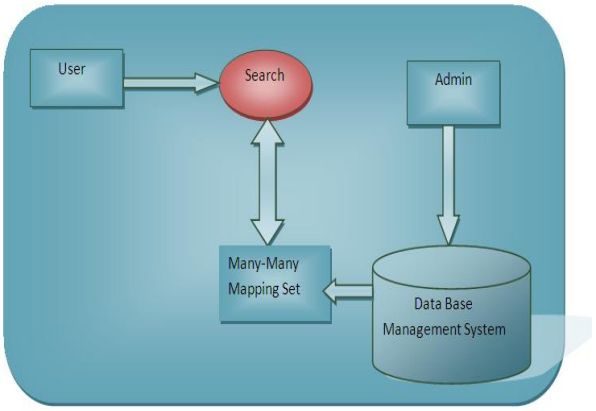
In CoRE multiple RelQuery is formulated based on the query entity α in conjunction with one or more RelTerms.

III. RELQUERY FORMULATION

For each query entity α, there are many possible formulations of a RelQuery, each of which is based on α and a conjunction of RelTerms. In particular, if n RelTerms are learned, then there are $(n^2 - 2)$ different combinations of RelTerms, leading to $(n^2 - 2)$ different formulation of RelQueries for each α.

To achieve this following techniques are used. 1) a confidence-aware termination condition, which estimates the confidence that a candidate target entity βc is the correct target entity and 2) a tree-based query formulation method, which selects a small subset of RelQueries to be issued as well as schedules the order of issuing those RelQueries .

IV. EXTENSION FOR MANY-TO-MANY MAPPING



System Architecture

Administrator

Administrator defines a set of tasks to perform and assigns role the user is allowed to perform.

Only authenticated person is allowed to get information or create the task.

Core Relation Creation

Context-Aware Relation Extraction method (CoRE) is particularly designed for the RC (Relation Completion) task.

For eg it creates core related information like colleges and its department, book, author and related information to store in the databases.

CoRE searches the web for documents that contain each of the seed instance pairs and from those documents it learns some “department” and “faculty”. It extracts a high-confidence target entity.

Many-to-Many Mapping Set

It automatically decides the number of target entities for each query entity. It describes relation completion between the source and target lists.

The following algorithm is used for many-to-many mapping technique:

The Naive string-matching algorithm

The naive algorithm finds all valid shifts using a loop that checks the condition $P[1 \ldots m] = T[s+1 \ldots s+m]$ for each of the $n - m + 1$ possible values of s.

```
NAIVE-STRING-MATCHER(T, P)
1  n ← length[T]
2  m ← length[P]
3  for s ← 0 to n − m
4       do if P[1..m] = T[s + 1..s + m]
5            then print "Pattern occurs with shift" s
```

Procedure NAIVE-STRING-MATCHER takes time $O((n - m + 1)m)$, and this bound is tight in the worst case.

The Procedure works as follows:

P is the source string provided by the user and T is the target string returned after executing the procedure [5].

The length of the source string (n) and target string (m) are identified and the loop is executed till n-m times.

Each letter in the source string is compared with the data in the database. The search is first conducted to find the accurate match. If it is found then the target string is the accurate match returned.

If accurate match is not found then the search is conducted to find the nearest matching string. If no match is found then an alert message is returned to the user.

V. CONCLUSION

By following many-to-many mapping technique the number of issues that arises in the search query are reduced.

A high confidence target entity is achieved. Effectiveness and efficiency of our techniques and models are improved and provides more flexibility and maintains high accuracy in RC task.

REFERENCE

[1] R. Gummadi, A. Khulbe, A. Kalavagattu, S. Salvi, and S. ambhampati, "Smartint: Using Mined Attribute Dependencies to Integrate Fragmented Web Databases," J. Intelligent Information Systems

[2] G. Koutrika, "Entity Reconstruction: Putting the Pieces of the Puzzle Back Together," HP Labs, Palo Alto, 2012.

[3] S. Chaudhuri, "What Next?: A Half-Dozen Data Management Research Goals for Big Data and the Cloud"

[4] Z. Li, M.A. Sharaf, L. Sitbon, S. Sadiq, M. Indulska, and X. Zhou, "Webput: Efficient Web-Based Data Imputation," Proc. 13th Int'l Conf. Web Information Systems Eng. (WISE).

[5] Introduction to Algorithms by Thomas H. Cormen, Charles E. Leiserson, and Ronald L. Rivest.