The Query Construction Interface for Analogous Information Retrieval
Radhika S. Mundhada¹ Prof. Dr.P.M.Jawandhiya²
¹Student ²Principal
¹,²PLITMS, Buldhana

Abstract— Today's web search engines, such as Google, MSN, yahoo, and amazon are generally work on a keyword search interface and a variety of statistical methods to catch users’ information needs. This query construction approach has been highly successful, as it is proven to be very intuitive and easy even for native web users. However, keyword search lacks the expressiveness to make use of the rich semantics in the Semantic Web. In order to construct a valid query to retrieve desired information, a user has to not only grasp the query language, but also understand the ontology thoroughly so as to find the right concepts and structures to form the query. Such a query construction process is complex and laborious. The work of the Semantics on Web is to equip the web with machine-processable and machine-understandable semantics, so that the Web can become a universal medium of information and knowledge exchange for both computer and human being. It can be foreseen that in the near future a lot of information on the Web will be described or annotated using these similar word, resulting in a huge knowledge base. Question arises how to utilize such rich and highly related words to improve the information seeking on the Semantic Web is becoming increasingly important. The Keyword Search approach makes it easy and the freely usable at the price of query expressiveness. This paper presents IQP—a novel approach to bridge the gap between usability of keyword search and expressiveness of database queries. IQP enables a user to start with an arbitrary keyword query and incrementally refine it into a structured query through an interactive interface.

Key words: IQP Incremental Query Plan, Semantic Web

I. INTRODUCTION

Now a day’s database are large and also the complex one which consisting of many entities. The users are interested to retrieve the information from the database by using SQL queries but the entire user is not familiar with SQL knowledge. Instead of traditional approach the Keyword Search makes it easy and the freely usable at the price of query expressiveness. In most of the cases these keyword search algorithms are also have no difference between the possible expected information needed by representation may receive adequate results. Databases enable users to precisely express their informational needs using structured queries.

Keyword search, originally developed for retrieving unstructured documents, is an intuitive and convenient interface for accessing unstructured data. However, since keyword search ignores most structured information, it may fail to understand the user’s intent and to identify the appropriate information. On the other hand, database queries allow users to exploit available structures to achieve more precise queries and corresponding result sets.

However, using a database system requires adequate knowledge of the database schema and proficiency of the query language, making data access a difficult task for unskilled users. Even for professionals, creating a database query that exactly matches the existing schema can be a laborious and error-prone process, especially when accessing a large and complex database. Today’s heterogeneous data management environments demand search interfaces that are not only sufficiently expressive to exploit structured queries, but also as intuitive and easy to use as keyword search. Structured queries are a powerful tool to precisely describe a user’s informational need and retrieve the intended information from a database. This task requires exact knowledge of the database schema as well as proficiency in a query language, which are typically beyond the expertise of end users.

Keyword search on the other hand can be performed efficiently by novice users, as it requires neither a-priori schema knowledge nor query construction skills. However, keyword search lacks expressiveness to precisely describe a user’s informational need, and may return irrelevant or incomplete results. To take advantage of both, i.e., expressiveness of structured queries and usability of keyword search, some recent approaches [1], [2],[4],[6] translate a keyword query into a ranked list of structured queries, such that the user can select the query that best represents her informational need. Such a ranking approach has two limitations.

1) As each keyword can potentially occur in any textual attribute of a database, the number of structural interpretations grows sharply with the complexity of the database schema and the length of the keyword query. With a complex database and a lengthy keyword query, it is computationally infeasible to materialize and sort all possible structural query interpretations at runtime.

2) Even a theoretically optimal ranking algorithm can, at best, rank the most common query interpretations highest, whereas the users with less frequent informational needs may not receive adequate results.

For example, if the majority of users who issued the keyword query “London” were interested in a city guide of London, the results referring to Jack London as a book author will receive a low rank. If a ranking function fails to place the user intended structured query within the top results, the user will need to examine all interpretations prior to the intended one. This process is tedious and error prone.

II. LITERATURE REVIEW & RELATED WORK

A. Authors: Barbara Poblete, Ricardo Baeza-Yates

- Title: Using Implicit Feedback and Query Patterns to Organize Web Documents
- Proposed Methodology:
  1) Presented a new document representation model based on implicit user feedback obtained from search engine queries
  2) Discovered the motivations of users when visiting a certain document
- Limitations:

All rights reserved by www.ijsrd.com 888
**Proposed Methodology:**

1) Their system is implemented in a 3-tier architecture, consisting of User Interface, Server and Database.

2) The experimental results show that IQP is highly helpful when user intended structured queries cannot be found within the top ranked results.

**Authors: X. Wang, C.X Zhai**

- **Title:** Learn from Web Search Logs to Organize Search Result
- **Journal & Publication Year:** ACM, 2007

**Proposed Methodology:**

1) Learnt “interesting aspects” of a topic from Web search logs and organized search results accordingly

2) Generated more meaningful cluster labels using past query words entered by users

**Limitations:**

1) Their method does not work if we try to discover user search goals of one single query in the query cluster rather than a cluster of similar queries

**Authors: Pavithra.G, Prabakar.D, Dr. Karthick.S**

- **Title:** A Survey on Query and Keyword Search in Database
- **Journal & Publication Year:** (IJSETR) Volume 2, Issue 1, January 2013

**Proposed Methodology:**

1) By using keyword search the results are produced but it may irrelevant to the user’s need, so with keywords the incremental query construction is discussed.

2) For searching the information from the database is analyzed, instead of using structured queries the keyword is used with incremental query construction

### III. ANALYSIS OF PROBLEM

Information retrieval by searching information on the web is not a fresh idea but has different challenges when it is compared to general information retrieval. Different search engines return different search results due to the variation in indexing and search process. The search engines only searches the information which is given on the web pages, recently, some research group’s start delivering results from their semantics based search engines, and however most of them are in their initial stages. Till none of the search engines come to close indexing the entire web content, much less the entire Internet. Current web is the biggest global database that lacks the existence of a semantic structure and hence it makes difficult for the machine to understand the information provided by the user. When the information was distributed in web, we have two kinds of research problems in search engine i.e.

1) How can a search engine map a query to documents where information is available but does not retrieve in intelligent and meaning full information?

2) The query results produced by search engines are distributed across different documents that may be connected with hyperlink. How search engine can recognize efficiently such a distributed results?

Most of the search engines search for keywords to answer the queries from users. The search engines usually search web pages for the required information. However they
filter the pages from searching unnecessary pages by using advanced algorithms. These search engines can answer topic-wise queries efficiently and effectively by developing state-of-art algorithms. However they are vulnerable in answering intelligent queries from the user due to the dependence of their results on information available in web pages.

The main focus of these search engines is solving these queries with close to accurate results in small time using much researched algorithms. However, it shows that such search engines are vulnerable in answering intelligent queries using this approach. They either show inaccurate results with this approach or show accurate but (could be) unreliable results. With the keywords based searches they usually provide results from blogs (if available) or other discussion boards.

By facing those points I will develop a system for incremental query construction using the keywords which are given by the user. In any system, searching is always a major task which requires careful understanding of user’s input and then providing the proper output requires careful understanding of the user’s query and his inputs. To do this we are designing a system for keyword recommendation so that the user can provide some words to the search engine and our system creates a query from those words. Then this query is given as an input to the search engine and an output is obtained, we then compare the accuracy of our system by comparing the outputs from both of these systems (one with our algorithm and one without our algorithm). Our system will give better and more accurate results as compared to the system without the recommendation engine.

IV. PROPOSED WORK

A structured query for the Semantic Web is composed of multiple concepts, properties and literals. The construction process of the structured query can be modeled as a hierarchy of query components. At the bottom of the hierarchy is the smallest component, where each is comprised of a single concept, a single property and a single keyword. The higher is the hierarchy more complex is the query components. Structured query lets users start with the smallest query components, and gradually evolve them into larger query components by climbing up the query hierarchy. For example, to search for the movie “Random Hearts”, a user might issue a keyword query “random crash Alcee”. For each of the terms, structured query provides a list of term-property combinations. For example, the user can specify whether “Alcee” should appear in the actor name, character name or movie title.

After the user specifies some basic query components, the system offers large components by climbing up the query hierarchy. For example, if the user specifies the character name “Alcee” and movie title “Random”, the system offers larger components that contain the selected smaller ones. For instance, after the user specifies the character name “Alcee” and movie title “Random”, the system can suggest the query component that connects these two term-property combination using the “actsin” property, as shown in the middle left of Figure afterwards. The user specifies that “crash” should appear in the plot-text, and the system can suggest the query component at the top, which is already the complete structured query required by the user. Usually, a user does not need to go through the complete construction process.

V. DESIRED IMPLICATIONS

Keyword search approaches gives the various modified results by applying various techniques such as,

1) The working of query construction program will work on the API approach and it gives to the NLP query language and by using this NLP language the keyword entered by a user get linked with the another query and this query is modified by the new keyword.

2) User first entered the query to search engine as in any form to request by user this search engine results are fetched by the API Application program Interface. This API helps to user for searching the results and then search click by user in search results.

3) After the clicking by user on search engine results gives to the NLP Natural Language Programming is applied to the search results to get the modified keyword.

4) This modified keyword gets linked with the query which is given to the user from the search results and this query is modified by new keyword.

5) This modified keyword is now click by user and this modified keyword is now shown to user for example user entered “apple green” to search engine the this time search engine reads all possible datasets which are in the database and this results to all possible query which contains an keyword “Apple”, such as “Apple green”, “Apple Eye”, “Apple Company”, “Apple Image” and so on.

6) This time the search engine gives lots of query formation which contain an keyword apple and the user have to click on the any one of the selected query such as “Apple Green” and this keyword get linked with the new keyword and the search engine now works on the selected query and this selected query then gives to search API.

7) The search API is applied to NLP and the modified keyword which contains various Adjectives, Nouns, Pronouns, and get sorted by those NLP language and the database is updated to get the requested output.

8) The process is continues in loop up to the user don’t get satisfied with the requested result. This is the mechanism which helps user to get satisfied in all manners in all drastic ways.

Keyword search on structured data has been extensively investigated in recent years. In [2],[3] the authors proposed mechanisms for conducting keyword search on the Semantic Web. Their approaches aim to predict structured
queries that best match user’s intents behind their keyword queries. However, as the number of possible structured queries grows exponentially with the size of RDF schema and the size of keyword queries, their approaches are only applicable to small datasets. Semantic Web goes beyond the state of the art, by using an incremental query construction process, so that it can be used on much larger data sources. In Structured query, we also devised efficient scheme for query optimization, so that query construction and query processing can be performed in reasonable response time.

The main contributions of this project are:
1) It will forward the problem of translating keywords into formal logic based queries for semantic search;
2) It will present a novel solution to this problem which is implemented as structured query that translates keywords into probabilistic queries;
3) It will provides a probabilistic query ranking model for picking most likely translated formal queries from keywords.

REFERENCES


[2] S.Durga Malleswari1, K.Kriran Kumar2, “Incremental Query Processing Based on Top K Results” IJDCST @October Issue. V-1, I-6, SW-16


[5] Xuanhui Wang, ChengXiang Zhai, ”Learn from Web Search Logs to Organize Search Results” Department of Computer Science University of Illinois at Urbana-Champaign Urbana, IL 61801 xwang20@cs.uiuc.edu


