*International Journal of Computer Systems (ISSN: 2394-1065), Volume 02– Issue 04, April, 2015*
*Available at http://www.ijcsonline.com/*

# Keyword Search on Query Routing Process

Janapati. Kalyani, Sri. K. Sreenivasulu

Department of Computer Science & Technology,
Madina Engineering College,
Andhra Pradesh, Kadapa, India.

***Abstract***

*Keyword search is associate intuitive paradigm for looking out connected knowledge sources on the net. It has a tendency to propose to route keywords solely to relevant sources to scale back the high value of process keyword search queries over all sources. It has a tendency to propose a completely unique methodology for computing top-k routing plans supported their potentials to contain results for a given keyword question. It has a tendency to use a keyword-element relationship outline that succinctly represents relationships between keywords and therefore the knowledge parts mentioning them. A structure grading mechanism is projected for computing the relevancy of routing plans supported scores at the amount of keywords, knowledge parts, component sets, and sub-graphs that connect these parts. Experiments dole out mistreatment a hundred and fifty public-ally accessible sources on the net showed that valid plans that square measure extremely relevant may be computed in one second on the average on one laptop. Further, we have a tendency to show routing greatly helps to boost the performance of keyword search, while not compromising its result quality.*

***Keywords:*** *Keyword search, keyword query routing (KDR), graph-structured data (GSD), RDF.*

## I. INTRODUCTION

In recent years the online has evolved from a world info area of joined documents to one wherever each documents and information are joined. Underpinning this evolution may be a set of best practices for publication and connecting structured information on the online called joined information. The adoption of the joined information best practices has because the extension of the online with a world information area connecting information from various domains like folks, companies, books, scientific publications, films, music, TV and radio programs, genes, proteins, medication and clinical trials, on-line communities, applied mathematics and scientific information, and reviews. This internet of information permits new styles of applications. There are generic joined information browsers which permit users to start out browsing in one information supply so navigate on links into connected information sources. There are joined information search engines that crawl the online by following links between information sources and supply communicative question capabilities over collective data, like however neighbourhood info is queried nowadays. The online of information conjointly unveil new potentialities for domain-specific applications not like internet two mishaps that work against a set of information sources, joined information applications treat high of associate unbound, international information area. This permits them to deliver a lot of complete answers as new information sources seem on the online.

It has a tendency to propose to research the matter of keyword question routing for keyword search over an oversized variety of structured and joined information sources. Routing keywords solely to relevant sources will scale back the high price of looking for structured results that span multiple sources. To the most effective of our information, the work given during this paper represents the primary arrange to address this downside.

It has a tendency to use a graph-based information model to characterize individual information sources. Therein model, it has a tendency to distinguish between associate element-level information graph representing relationships between individual information components, and a set-level information graph, that captures info regarding cluster of components. This set-level graph basically captures a locality of the joined information schema on the online that's depicted in RDFS, i.e., relations between categories. Often, a schema may be incomplete or just doesn't exist for RDF information on the online. In such a case, a pseudo schema is often obtained by computing a structural outline like a data guide.

## II. PROPOSED APPROACH

In this project propose to route keywords solely to relevant sources to scale back the high value of process keyword search queries over all sources. It tends to propose a completely unique technique for computing top-k routing plans supported their potentials to contain results for a given keyword question and use a keyword-element relationship outline that succinctly represents relationships between keywords and therefore the information components mentioning them.

## III. KEYWORD SEARCH

Keyword analysis could be a follow utilized by computer program optimization professionals to seek out

and analysis actual search terms individuals enter into the search engines once conducting an inquiry. Computer program optimization professional's analysis keywords so as to realize higher rankings in search engines. Once a distinct segment keyword is found, it's expanded upon to seek out similar keywords. The method is sometimes assisted by keyword suggestion tools, which provide synonym finder and alternate keyword suggestion practicality. Most of the time the varied search engines give their own keyword suggestion tools moreover that additionally embrace the quantity of searches created for every of these keywords. This info is then employed in order to pick the proper keyword counting on the SEO goals of the web site. Existing work on keyword search relies on an element-level model (i.e., data graphs) to compute keyword query results.

Existing work are often categorized into two main categories:

>schema-based approaches

>Schema-agnostic approaches

There is a unit schema-based approach enforced on prime of off-the-rack databases. A keyword question is processed by mapping keywords to components of the info (called keyword elements). Then, mistreatment the schema, valid be part of sequences area unit derived, that area unit then used to affix ("connect") the computed keyword components to create therefore known as candidate networks representing potential results to the keyword question.

Schema-agnostic approaches operate directly on the information. Structured results area unit computed by exploring the underlying knowledge graph. The goal is to search out structures within the knowledge known as Steiner trees (Steiner graphs in general), that connect keyword components. Varied types of algorithms are projected for the economical exploration of keyword search results over knowledge graphs, which could be terribly massive.

## IV. KDR

A Query Routing Model: a technique that finds the most suitable person to help you based on knowledge, trust and activity. To materialize an intention-based retrieval system we need to extend the traditional IR paradigm with respect to three crucial technological components:

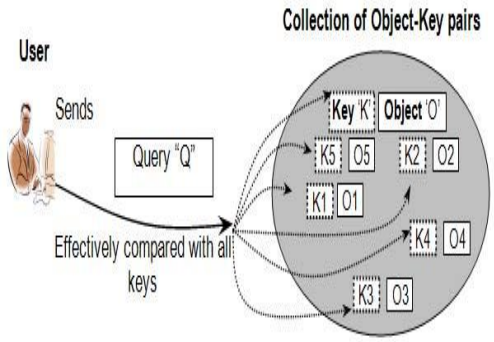



Fig 1: The search paradigm

1. The key which is used to retrieve objects (items).
2. Comparison of keys.
3. The index, which reduces the actual number of comparisons required during the search.

In the traditional IR paradigm, a database is a collection of searchable objects and key pairs. Generally the key is a number, text string, a set or vector of keywords which represent a compact description of an object. The query has a structure which is similar to that of the object key (set, vector, etc.), and it is effectively compared against all object keys to identify the matching objects using the index. To extend this paradigm for intention-based searching, the key should be a "semantic descriptor" data structure that represents the meaning of user's intention and the object. The key comparison should ascertain the similarity between meanings represented by two semantic descriptors, whereas the index should be a scalable infrastructure designed to improve the search speed and recall.

**ALGORITHM:**

```
if (expansion == TRUE)
  q = synonyms (q,DEFAULT_AMOUNT)

for each f_u in Followers_u do
  Begin
    k_{f_u,q} = knowledge (f_u,q)
    t_{u,f_u} = similarity (f_u,u) * friendship (f_u,q)
    a_{f_u}   = activity (f_u)
    f_u.victories = 0
  End

/*Weight Product Model*/
for each f1_u in Followers_u do
  for each f2_u in Followers_u do
    Begin
      if (f1_u == f2_u)
        continue;

      k = k_{f1_u,q} + k_{f2_u,q}
      t = t_{u,f1_u} + t_{u,f2_u}
      a = a_{f1} + a_{f2}

      comparison = (k^RELEVANCE_KNOWLEDGE_LEVEL) *
                   (t^RELEVANCE_TRUST_LEVEL) *
                   (a^RELEVANCE_ACTIVITY_LEVEL)
      if (comparison~ = 1)
        Begin
          f1_u.victories ++
          f2_u.victories ++
        End
      else if (comparison~ > 1)
        f1_u.victories ++
      else if (comparison < ~1)
        f2_u.victories ++
    End

sort_according_victories (Followers_u)

return Followers_u
```

## V. GSD

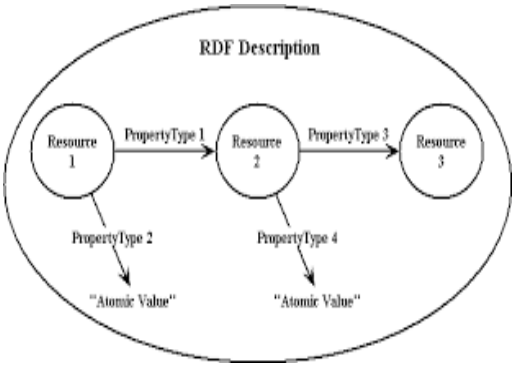
The standard graph of a social network has several people and attached to them content items identified by numbers which are supposed to be time stamps. The ego network changed from a star topology to a list topology and each ego network has a certain edge type which is modeled by edge color here. This graph stores exactly the same information as the standard model but makes retrieval of news streams much faster

## VI. RDF

The Resource Description Framework (RDF) is a family of World Wide Web Consortium (W3C) specifications originally designed as a metadata data model. It has come to be used as a general method for conceptual description or modeling of information that is implemented in web resources, using a variety of syntax notations and data serialization formats. It is also used in knowledge management applications.

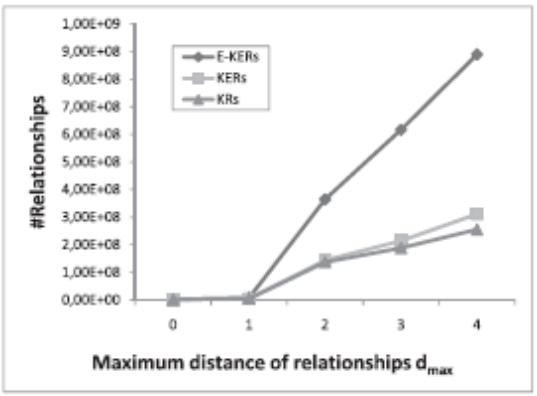
The RDF data model[2] is similar to classical conceptual modeling approaches such as entity–relationship or class diagrams, as it is based upon the idea of making statements about resources (in particular web resources) in the form of subject–predicate–object expressions. These expressions are known as triples in RDF terminology. The subject denotes the resource, and the predicate denotes traits or aspects of the resource and expresses a relationship between the subject and the object. For example, one way to represent the notion "The sky has the color blue" in RDF is as the triple: a subject denoting "the sky", a predicate denoting "has", and an object denoting "the color blue". Therefore RDF swaps object for subject that would be used in the classical notation of an entity–attribute–value model within object-oriented design; object (sky), attribute (color) and value (blue). RDF is an abstract model with several serialization formats (i.e., file formats), and so the particular way in which a resource or triple is encoded varies from format to format.

This mechanism for describing resources is a major component in the W3C's Semantic Web activity: an evolutionary stage of the World Wide Web in which automated software can store, exchange, and use machine-readable information distributed throughout the Web, in turn enabling users to deal with the information with greater efficiency and certainty. RDF's simple data model and ability to model disparate, abstract concepts has also led to its increasing use in knowledge management applications unrelated to Semantic Web activity.

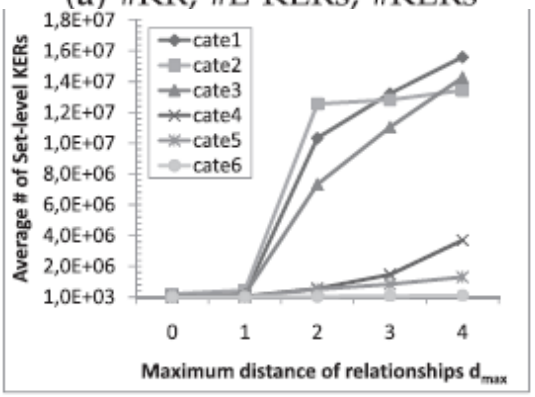



A collection of RDF statements intrinsically represents a labeled, directed multi-graph. As such, an RDF-based data model is more naturally suited to certain kinds of knowledge representation than the relational model and other ontological models. However, in practice, RDF data is often persisted in relational database or native representations also called Triple stores, or Quad stores if context (i.e. the named graph) is also persisted for each RDF triple.ShEX, or Shape Expressions, is a language for expressing constraints on RDF graphs. It includes the cardinality constraints from OSLC Resource Shapes and Dublin Core Description Set Profiles as well as logical connectives for disjunction and polymorphism. As RDFS and OWL demonstrate, one can build additional ontology languages upon RDF.

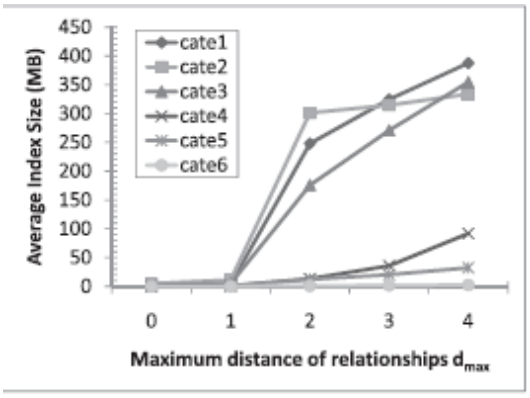
## VII. RESULT ANALYSIS



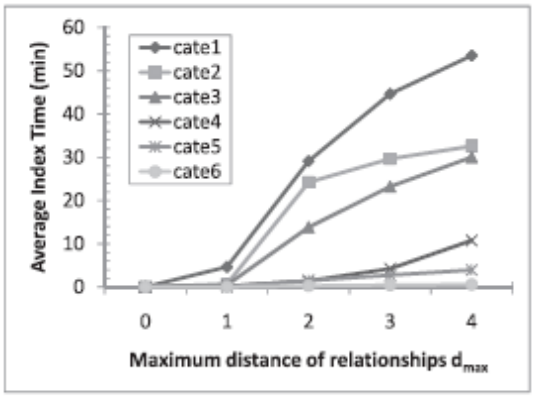

(a) #KR, #E-KERs, #KERs




(b) #KERs




(c) KERG size




(d) KERG building time

## VIII. CONCLUSION & FUTURE WORK

A conferred a solution to the novel drawback of keyword question routing. Supported modeling the search space as a structure inter-relationship graph, we tend to projected an outline model that teams keyword and element relationships at the amount of sets, and developed a structure ranking theme to include connectedness at different dimensions. The experiments showed that the summary model succinctly preserves relevant data. In combination with the projected ranking, valid plans (precision@1 ¼ 0:92) that square measure extremely relevant (mean reciprocal rank ¼ 0:86) might be computed in one's on the average. Further, we tend to show that once routing is applied to associate degree existing keyword search system to prune sources; substantial performance gain may be achieved.

## REFERENCES

[1] V. Hristidis, L. Gravano, and Y. Papakonstantinou, "EfficientIR-Style Keyword Search over Relational Databases," Proc. 29th Int'l Conf. Very Large Data Bases (VLDB), pp. 850-861, 2003.

[2] F. Liu, C.T. Yu, W. Meng, and A. Chowdhury, "Effective Keyword Search in Relational Databases," Proc. ACM SIGMOD Conf., pp. 563-574, 2006.

[3] Y. Luo, X. Lin, W. Wang, and X. Zhou, "Spark: Top-K Keyword Query in Relational Databases," Proc. ACM SIGMOD Conf., pp. 115-126, 2007.

[4] M. Sayyadian, H. LeKhac, A. Doan, and L. Gravano, "Efficient Keyword Search Across Heterogeneous Relational Databases," Proc. IEEE 23rd Int'l Conf. Data Eng. (ICDE), pp. 346-355, 2007.

[5] B. Ding, J.X. Yu, S. Wang, L. Qin, X. Zhang, and X. Lin, "Finding Top-K Min-Cost Connected Trees in Databases," Proc. IEEE 23rd Int'l Conf. Data Eng. (ICDE), pp. 836-845, 2007.

[6] B. Yu, G. Li, K.R. Sollins, and A.K.H. Tung, "Effective Keyword-Based Selection of Relational Databases," Proc. ACM SIGMOD Conf., pp. 139-150, 2007.

[7] Q.H. Vu, B.C. Ooi, D. Papadias, and A.K.H. Tung, "A Graph Method for Keyword-Based Selection of the Top-K Databases," Proc. ACM SIGMOD Conf., pp. 915-926, 2008.

[8] V. Hristidis and Y. Papakonstantinou, "Discover: Keyword Search in Relational Databases," Proc. 28th Int'l Conf. Very Large Data Bases (VLDB), pp. 670-681, 2002.

[9] L. Qin, J.X. Yu, and L. Chang, "Keyword Search in Databases: The Power of RDBMS," Proc. ACM SIGMOD Conf., pp. 681-694, 2009.

[10] G. Li, S. Ji, C. Li, and J. Feng, "Efficient Type-Ahead Search on Relational Data: A Tastier Approach," Proc. ACM SIGMOD Conf., pp. 695-706, 2009.

[11] V. Kacholia, S. Pandit, S. Chakrabarti, S. Sudarshan, R. Desai, and H. Karambelkar, "Bidirectional Expansion for Keyword Search on Graph Databases," Proc. 31st Int'l Conf. Very Large Data Bases (VLDB), pp. 505-516, 2005.

[12] H. He, H. Wang, J. Yang, and P.S. Yu, "Blinks: Ranked Keyword Searches on Graphs," Proc. ACM SIGMOD Conf., pp. 305-316, 2007.