**Available Online at www.ijcsmc.com**

**International Journal of Computer Science and Mobile Computing**

**A Monthly Journal of Computer Science and Information Technology**

**ISSN 2320–088X**

*IJCSMC, Vol. 3, Issue. 1, January 2014, pg.465 – 469*

**SURVEY ARTICLE**

# A SURVEY OF QUESTION ANSWERING SYSTEM FOR COMPARABLE ENTITY MINING

**S.Deepa[1], S.Vasumathi Kannagi[2]**

[1]P.G Scholar, Computer Science and Engineering &Anna University, India
[2]Assistant Professor, Computer Science and Engineering &Anna University, India
[1] Sdeepa153@gmail.com; [2] Vasumathi.arun@gmail.com

***Abstract- Comparing individual object with one more is a characteristic component of individual assessment making process. Though, it is not forever straightforward to be familiar with what to evaluate and what are the alternating entity to find the exact information retrieval system. In this paper, learning the number of technique that mechanically evaluates the questions and entity designed for mining consumer outcome for information extraction. Several methods have been anticipated in earlier work to solve the problem of entity mining outcomes and none of the work have been studied the accurate foundation of comparable entities, but supervised knowledge technique give an well-organized learning outcomes for comparable entities for different questions and routinely mine equivalent entities as of relative query with the intention of users posted online. In this article we study the difficulty of earlier work query and answering scheme and finish why we preferring equivalent entity mining.***

**Index terms:** Entity mining; Question answering system; Question Answering Challenge (QAC); Machine learning methods such as Support Vector Machines (SVM); Bayesian learning. Sequential rules (CSRs); label sequential rules (LSRs).

## I. INTRODUCTION

With the sudden increasing of usage of information or data from web pages or WWW in Internet, it is important to find the similar entities to mine the result exactly , Natural foreign language Question Answering system is documented as a ability through enormous prospective. Conservatively, QA have be used by frequent Artificial Intelligence (AI) investigators, except the majority of Question Answering (QA) system have been developed be games restricted in the direction of lab and an extremely constrained area. Further in recent times have been planned a Question Answering (QA) system go after to stimulate the investigation intended for real world purpose.

Different patterns based Question Answering (QA) System with the purpose of make inquiries concerning description by Harabagiu Sanda et.al [1]. This variety of example related to dissimilar patterns was expansively second-hand through our Question Answering (QA) scheme, even though in numerous belongings they are effective in grouping through various added displays. It is moreover important and did not imprison the make use of of these example to question interested concerning

 

description. Then presume, in universal, with the intention of present be supposed to not exist one-to-one association among a specified pattern and a query category.

The similar pattern be capable of be suitable in answering frequent group of questions. One of the majorities significant behaviors of estimating an entity being in the direction of straightforwardly match up to it through a comparable entity. The purpose of this effort is to mine and to examine relative entity in information on the Web and web log files. This assignment has numerous significant appliances. For instance, following an original web pattern for similar question with same entity is launched and another similar user is compared with individuals of its opponent. Extracting such type of patterns with entity can help out commerce in its advertising and creation benchmarking hard work.

In the WWW period, an assessment action characteristically involves the following ways: investigation of significant WebPages surrounding information concerning the final result of the information retrieval, discover user result with same entity and recognize final results. In this survey paper, focal point on discovery a group of equivalent entities specified a user's contribution entity.

First learning the difficulty of relative entity mining. It has two majorities of important tasks:

1. Specified a set of assessment entity, categorize relative entities beginning them, and categorize the recognized relative entity addicted to dissimilar types of entity from original web information or data retrieval.

2. Extract family member entity relations beginning the recognized information from web for each and every user. This contains the extraction of web information establishment entities and their matching features that related to user searching results with the intention of being compared, and relative keywords

In this paper study the current approach designed for routinely knowledge such expected language alongside through formative their correctness beginning then us information consequences, designed for specified categories of questions. Our techniques make use of the machine knowledge procedure of bootstrapping to construct a huge labeled quantity preliminary by means of simply a small number of pattern of Question Answering (QA) pairs. Must note down to every search engines agree to a consumer to go through a normal foreign language query as an alternative of an inquiry.
The search engines subsequently eliminate assured regular stop words and delight the respite of the query as a question. Accordingly search engines perform because qualification they are able to handle the primary phase of contribution in the variety of a question. Conversely, they still make obtainable documents somewhat than response as their yield outcome.

In the following section compare the dissimilar methods is used for comparable entity mining ,section 2 compare the different methods and problems of every methods ,section 3 defines the problem of the earlier work and solution of the problem ,finally concludes the survey result ,how to solve the problems direction (future work also given).

## II. LITERATURE SURVEY

In the comparable entity mining result they are third categories of analysis performed, first the primary questions are analysis to identify different results, second sentence extraction methods are studied and their problems also, finally machine learning methods to rank the results are studied with their individual problems of each and every methods.

### 2.1. Question analysis

In web search engines the detection of query keywords suggested by user and recommend conventional answer category becomes an significant for Question and Answering scheme to extract web information exactly for each and every user .Even though the majority of the schema depends on categorization of predictable answer category, the amount of nodes in the categorization differ extensively beginning only number to a small number of thousands.

Abney et al. [2] second-hand a dissimilar types of question and answered related procedures methods are all of the procedures depends on entity detection schema techniques. On behalf of every one query, a position of appropriate passageway with the intention of frequently include the answers is initially recognized. A candidate set consists of set of entity related information for every user and their similar paragraph are extracted from these entity specified by user. Together the query and these mined entities are categorized into a predetermined candidate set of classification results. Purely individual's entities with the purpose

 

of equivalent the categories essential by the question are preserved and categorized over by means of the occurrence and further situation associated information.

In expressions of determining associated substance intended for an entity, our effort is comparable to the examiner on recommender scheme, which suggest substance to a customer. A recommender system mostly depends on measuring the relationship among entity and their arithmetical association in consumer web log information [3]. For instance, Amazon commends goods to its regulars supported on their individual acquire history; comparable customers' acquire record, and relationship among yield. Although, recommending an item does not equivalent to judgment a similar item. In the case of Amazon, the purpose of suggestion is to attract their clients to attach more items to their shopping carts by suggestive of like or related items. Although in the case of assessment, would like to facilitate user explore alternative, i.e., helping them make a conclusion amongst equivalent items.

In the unique case, the TREC quantity is used as the key basis and Information Retrieval (IR) was achieved by the Information Retrieval (IR) factor of our Question Answering (QA) scheme [4] .In the subsequently case, the network be the contribution introduction position and the Information Retrieval (IR) was performed by the AltaVista search steam engine.

Harabagiu et al. [5] furthermore used WordNet to assist in answer description questions. Though, they take the hypernyms of the phrase to be defined as the evasion answers whereas we make used of its glosses. In this casing it may not advance the desired response to the top other than it might considered as a substitute "authenticate" as a good quality answer. More examine is required to explore the exchange among using hypernyms as well as glosses. WordNet glosses are integrated in IBM's numerical question answer system as explanation features.

### 2.2. Passage extraction

The major aim of this work is to provide a text pool based controllable size of schema for extract candidate answers. The popular peak performing scheme in TRECs makes use of their individual retrieval technique for each and every passages [6]. Final result of the schema also stored in database and measured by the quantity of each and every passage using TREC dataset with elevated precision attain also have to resolve for answers were every sentences matched as a effect of little precision pattern. The WHY-FAMOUS question kind is exclusion and probably determination is suitable to the information that the construction was tested on a diminutive number of questions.

Extract contestant answer corresponding to respond types. If the estimated answer categories are representative named entities, in sequence extraction engines [7] are used to extract applicant answers. Otherwise extraordinary answer pattern also used to every pin-point results. For instance, a group of 6 response patterns for explanation questions. Our job on comparator mining is correlated to the investigate on entity and relative extraction in to extract information from web pages [8]. Particularly, the for the most part relevant job [9], [10] on taking out relative sentence and relations. Their technique applied category CRS and LRS learned as of interpret corpora to classify comparative sentences and extract comparative relations correspondingly in the news and analysis area. The comparable technique can be helpful to relative question appreciation and comparator mining beginning questions. Jindal and Liu [9-10], which is second-hand as baseline for evaluation and characterize the well suited schema. In this work initially used rule based system such as CRS &LRS in their advance, and after that illustrate their comparative mining technique.

However, their method typically can get high precision but suffer from low recall [10]. However, ensure high recall is crucial in our intentional application scenario somewhere users can problem arbitrary queries. To address this problem, we expand a weakly organized bootstrapping model learn method by efficiently leveraging unlabeled questions.

Bootstrapping methods all well efficient method to retrieve information from web pages and compare every user efficiently [11]. Our work is comparable to them in conditions of method by means of bootstrapping procedure to mine entities by means of a precise relative. Though, our assignment is dissimilar beginning theirs is to facilitate it necessitate not simply extract individual entity except also ensure with the intention of the individual entity is extracted on or after relative question classification, which is normally not necessary in information extraction (IE) process.

In earlier work [12] and [13 ] used an learning methods like Naïve Bayesian model and Support vector machine (SVM) using the category sequential regulations as features. For evaluation principles, also experimentation which is measured to be individual of the best classifier construction techniques. In conclusion performed experimental assessment using three types of categories and Internet opportunity consideration. Our assignment is to categorize relative sentences in such key texts.

 

### 2.3. Answer ranking

Answer ranking assign attains to contestant answers corresponding to their regularity in peak ranked based paragraphs with individual similarity measure for individual candidate answers are extracted from exterior sources [5] & [6] among the dependence formation of questions and contestant answers and equivalent of predictable answer category.

Investigate scheme contribution in TRECs [14] and the coming QAC listening carefully on the difficulty answering each and every schema using the QAC to facilitate encloses small fact-based answers beginning a huge compilation of text.

Passage recovery procedure is applied for preliminary preprocessing [15]. The passageway ranking algorithm make use of semantic competition information among the opposite manuscript occurrence term, weighting data of every term, and moreover the exposure of these query related terms in the passageway itself. A numerical background free of claim sentence construction parser used with WordNet tool to establish the query class. Then these patterns are matched to query class, and the probable answers are extracted and graded using a variety of quality heuristics.

Learning search engine for user a comprehensive query modification for query responding system [16]. A comparable alteration procedure also become visible is with the intention of the existing query passage summit of the majority general search engines don't make available sufficient competence for straight query reply in a natural verbal communication approach. It can considerably progress the probability of decision good answers at the peak of the investigate of alteration rules are trained and functional to the query at explore time, and the experimentation includes two capable outcomes. Their effort, conversely, is listening carefully on improving the probability of receiving elevated quality papers beginning a investigate train. It might not make available some machine to recognize the accurate answers beginning the investigate outcomes.

In regulate to assist to extraction of answers from web information and build the categorization assignment easier; the query is categorized into small, arithmetical and sequential. It is like similar to methods used by [16] a variety of question inflection methods are applied to the preliminary query to acquire elevated superiority results for afterward answer extraction. Their respond extraction component makes use of together TF-IDF term frequency schema and distance based word measure to extract answers. Score values are assigned to extracted results based on their background in manuscript and subsequently supplementary grouped into similar groups. The component through the uppermost achieve in every cluster is chosen as a representative for with the intention of collection and is existing to the consumer as a possible answer.

Our scheme is dissimilar from [17] don't make use of a profound normal verbal communication parser. As the writer acknowledged on the web pages are extremely time-consuming and are not utilizable on the Web. It works similar to the working schema of parser in corresponding to formulate it scalable. As an alternative, substitute these times overwhelming parts by means of move toward based on rule-based classification systems and probabilistic expression reranking and at rest accomplish realistic management.

## III. CONCLUSION

In the writing discovering methods move toward effort to differentiate among the main and resulting query words. In this characteristic is conversed as functional to penetrating for result to Question Answering system, everywhere the answers are correspond to as individual entity . It establish the dissimilar knowledge techniques are second-hand to study the dissimilar equal entity join up by means of user specified keyword main keywords through comparator pairs with the intention of remove numerous information from extraction question patterns and rank the similar pairs by means of knowledge techniques with the intention of communicate the concentrate of every and all entity similar to a different user. They cannot be disregarded.

To conquer these problems solve the issues using learning methods, weakly supervised technique to recognize relative questions and mine comparator pairs concurrently. It depends on the explanation approaching that a high-quality relative question recognition pattern be supposed to extract good quality of comparators be supposed to happen in good quality of relative subjects to bootstrap the extraction and detection procedure.
In our prospect effort, arrangement to append disambiguation. The structure, get better extraction question pattern appliance result and extract exceptional question patterns. Superior tunes a variety of phrase of the scheme, and experimentation through added heuristics in answer assortment.

## REFERENCES

1) Harabagiu Sanda, Dan Moldovan, Marius Pasca, Rada Mihalcea, Mihai Surdeanu, Razvan unescu, Roxana Girju, Vasile Rus and Paul Morarescu. "Falcon: Boosting Knowledge for Answer Engines", Proceedings of the Text Retrieval Conference (TREC-9), 2000.

 

2) Abney, S., M. Collins, and A. Singhal,"Answer Extraction. In Proceedings of the Applied Natural Language Processing Conference (ANLP-NAACL-00), Seattle, WA, 296–301,2000.

3) S. Li, C.-Y. Lin, Y.-I. Song, and Z. Li, "Comparable Entity Mining from Comparative Questions," Proc. 48th Ann. Meeting of the Assoc. for Computational Linguistics (ACL '10), 2010.

4) Lin, C-Y," The Effectiveness of Dictionary and Web-Based Answer Reranking", Proceedings of the COLING-2002 conference. Taipei, Taiwan, 2002 .

5) Harabagiu, S., D. Moldovan, M. Pasca, R. Mihalcea, M. Surdeanu, R. Buneascu, R. Gîrju, V. Rus and P. Morarescu ,"FALCON: Boosting Knowledge for Answer Engines. Proceedings of the 9th Text Retrieval Conference (TREC-9), NIST, 479–488,2001.

6) Brill, E., J. Lin, M. Banko, S. Dumais, and A. Ng," Data-Intensive Question Answering", Proceedings of the TREC-10 Conference. NIST, Gaithersburg, MD, 183–189,2001.

7) Srihari, R. and W. Li," A Question Answering System Supported by Information Extraction", Proceedings of the 1st Meeting of the North American Chapter of the Association for Computational Linguistics (ANLPNAACL- 00), Seattle, WA, 166–172,2000.

8) E. Riloff and R. Jones, "Learning Dictionaries for Information Extraction by Multi-Level Bootstrapping," Proc. 16th Nat'l Conf. Artificial Intelligence and the 11th Innovative Applications of Artificial Intelligence Conf. (AAAI '99/IAAI '99), pp. 474-479, 1999.

9) N. Jindal and B. Liu, "Identifying Comparative Sentences in Text Documents," Proc. 29th Ann. Int'l ACM SIGIR Conf. Research and Development in Information Retrieval (SIGIR '06), pp. 244-251, 2006.

10) N. Jindal and B. Liu, "Mining Comparative Sentences and Relations," Proc. 21st Nat'l Conf. Artificial Intelligence (AAAI '06), 2006.

11) D. Ravichandran and E. Hovy, "Learning Surface Text Patterns for a Question Answering System," Proc. 40th Ann. Meeting on Assoc. for Computational Linguistics (ACL '02), pp. 41-47, 2002.

12) Liu, B. Web Data Mining: Exploring hyperlinks, Contents, and Usage Data. A forthcoming book. 2006/2007.

13) Joachims, T. Making large-scale SVM learning practical. Advances in Kernel Methods - Support Vector Learning, B. Schölkopf and C. Burges and A. Smola (ed.), 1999.

14) Fukumoto, J, T. Kato, and F. Masui," NTCIR Workshop 3 QA Task – Question Answering Challenge (QAC)",http://research.nii.ac.jp/ntcir/ workshop/qac/cfp-en.html,2001.

15) Clarke, C.L.A., Cormack, G.V., Kisman, D.I.E., & Lynam, T.R," Question answering by passage selection (multitext experiments for TREC-9)", In NIST Special Publication 500-249: The Ninth Text Retrieval Conference (TREC 9) (pp. 673–683). Maryland: NIST,2000.

16) Agichtein, E., Lawrence, S., & Gravano, L. ,"Learning search engine specific query transformations for question answering," In The Proceedings of the 10th World Wide Web Conference (WWW2001). New York: ACM, 2001.

17) Kwok, C., Etzioni, O., & Weld, D.S ,"Scaling question answering to the web", In Proceedings of the 10th World Wide Web Conference (WWW 2001).

 