

# Selecting Contextual Peripheral Information for Answer Presentation: The Need for Pragmatic Models

**Rivindu Perera, Parma Nand**

School of Computer and Mathematical Sciences

Auckland University of Technology

Auckland, New Zealand

{rivindu.perera, parma.nand}@aut.ac.nz

## Abstract

This paper explores the possibility of presenting additional contextual information as a method of answer presentation Question Answering. In particular the paper discusses the result of employing Bag of Words (BoW) and Bag of Concepts (BoC) models to retrieve contextual information from a Linked Data resource, DBpedia. DBpedia provides structured information on wide variety of entities in the form of triples. We utilize the QALD question sets consisting of a 100 instances in the training set and another 100 in the testing set. The questions are categorized into single entity and multiple entity questions based on the number of entities mentioned in the question. The results show that both BoW (syntactic models) and BoC (semantic models) are not capable enough to select contextual information for answer presentation. The results further reveals that pragmatic aspects, in particular, pragmatic intent and pragmatic inference play a crucial role in contextual information selection in the answer presentation.

## 1 Introduction

Answer Presentation is the final step in Question Answering (QA) which focuses on generating an answer which closely resemble with a human provided answer (Perera, 2012b; Perera, 2012a; Perera and Nand, 2014a). There is also a requirement to associate the answer with additional contextual information when presenting the answer.

This paper focus of exploring methods to extract additional contextual information to present with the

extracted factoid answer. We provide a classification if questions based on the type of the answer required and the number of entities that mentioned in the questions. The question classification is illustrated in Fig. 1. Firstly, question can be categorized based on the information need where questions may require a definition as the answer or a factoid answer which is an information unit (Perera, 2012a; Perera and Nand, 2014a). The definitional questions need definitions which include both direct and related background information and there is no need to further expand the answer with contextual information. So far the way factoid questions presentation involved only the answer itself without contextual information. Recently, Mendes and Coheur (2013) argued that even factoid questions need to present additional information. An advantage of presenting contextual information is that answer is justified by the information provided, so that users can conclude that the answer that is acquired by the system is one that they are searching for.

The rest of the paper is structured as follows. Section 2 explores BoW and BoC models to rank contextual information. Section 3 focuses on presenting the experimental framework. Section 5 presents information on related work and we conclude the paper in Section 6.

## 2 Content selection using weighted triples

This section presents models to rank triples focusing on open domain questions as communicative goals. Open domain questions require knowledge from different domains to be aggregated which making it more challenging compared to simply generating a

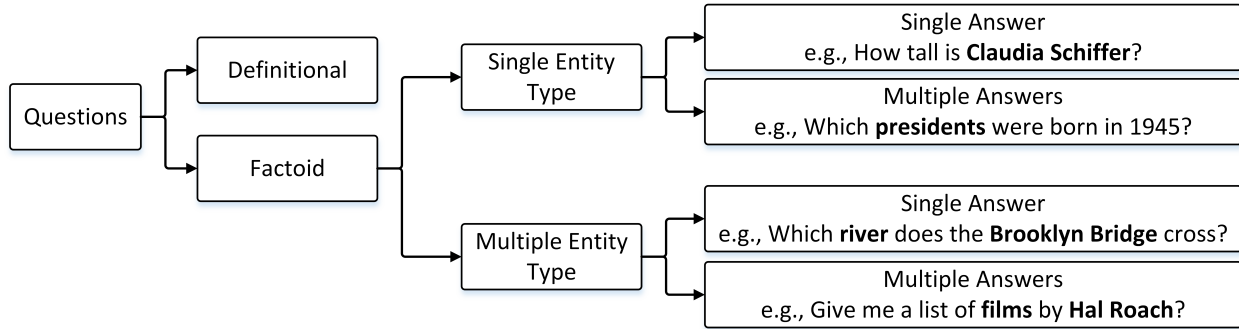


Figure 1: Classification of common question types

content for given single theme topic. Our objective is to select a set of triples which can be used to generate a more informative answer for a given question.

We investigate the problem from two perspectives; as a Bag of Words (BoW) and as a Bag of Concepts (BoC). In the following sections, we discuss the strategy used for ranking and the details on the supporting utilities including domain corpus, reference corpus, triple retrieval, and threshold based selection.

The high level design of the framework used to experiment BoW and BoC is shown in the Fig. 2. The model utilizes two corpora (domain and reference) and selectively used based on the requirement. The domain corpus is constructed using search snippets collected from the web by using information from the question and answer as query terms. The reference corpus represents knowledge about general domain. The model also has utility functions to retrieve triples using SPARQL queries, filter the duplicates, and to perform basic verbalization.

## 2.1 The problem as a Bag of Words

We utilized token similarity, Term Frequency - Inverse Document Frequency (TF-IDF), and Residual Inverse Document Frequency (RIDF) in two flavours which are widely used in information retrieval tasks. The following sections describes these models in detail.

### 2.1.1 Token similarity

Token similarity ranks the triples based on the appearance of the terms in triple and the question being considered. In particular we employ the cosine

similarity (1) to calculate the similarity between the tokenized and stopwords removed question/answer and the triple.

$$\begin{aligned}
 sim_{cosine}(\vec{Q}, \vec{T}) &= \frac{\vec{Q} \cdot \vec{T}}{|\vec{Q}| |\vec{T}|} \\
 &= \frac{\sum_{i=1}^n Q_i T_i}{\sqrt{\sum_{i=1}^n Q_i^2} \sqrt{\sum_{i=1}^n T_i^2}}
 \end{aligned} \tag{1}$$

Here, Q and T represent the question and the triple respectively.

### 2.1.2 Term Frequency – Inverse Document Frequency (TF-IDF)

The TF-IDF (2) is used to rank term ( $t$ ) from the question and answer present in the triple ( $T$ ). A triple is then associated with a weight which is the sum of the weights assigned to the triple terms.

$$\begin{aligned}
 TF - IDF(Q, T) &= \sum_{i \in Q, T} tf_i \cdot idf_i \\
 &= \sum_{i \in Q, T} tf_i \cdot \log_2 \frac{N}{df_i}
 \end{aligned} \tag{2}$$

Where  $tf$  represents the term frequency,  $N$  stands for number of documents in the collection and  $df$  is the number of documents with the corresponding term.  $Q$  represents the question, however in our experiment we tested the possibility of utilizing a domain corpus instead of the original question or the question with the answer.

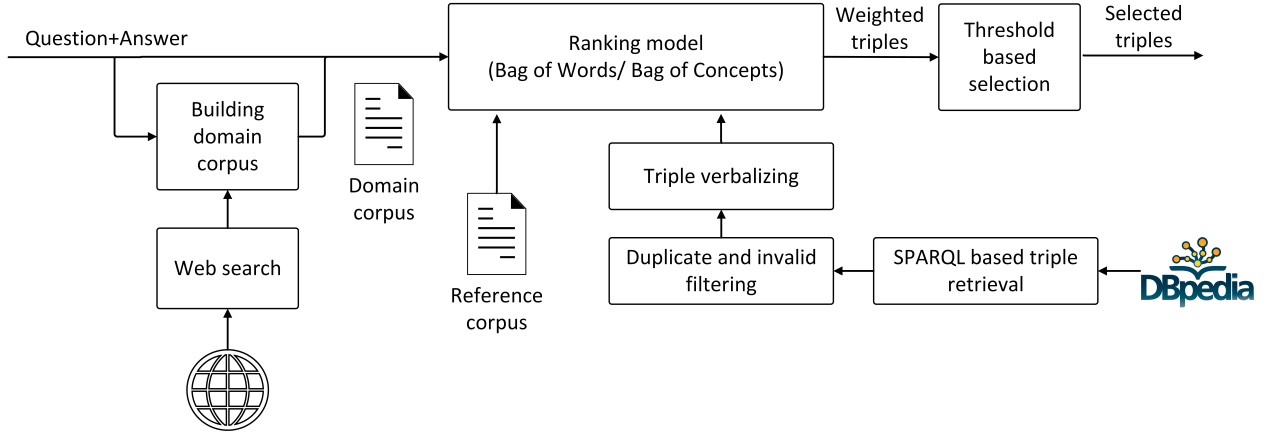


Figure 2: The schematic representation of the content selection framework

### 2.1.3 Okapi BM25

Okapi ranking is an extension to the TF-IDF that is based on the probabilistic retrieval framework.

The Okapi ranking function can be defined as follows:

$$Okapi(Q, T) = \sum_{i \in Q, T} \left[ \log \frac{N}{df_i} \right] \cdot \frac{(k_1 + 1) tf_{i,T}}{k_1 \left( (1 - b) + b \left( \frac{L_T}{L_{ave}} \right) \right) + tf_{i,T}} \cdot \frac{(k_3 + 1) tf_{i,Q}}{k_3 + tf_{i,Q}} \quad (3)$$

Where,  $L_T$  and  $L_{ave}$  represent the length of the triple and average of length of a triple respectively. The Okapi also uses set of parameters where  $b$  is usually set to 0.75 and  $k_1$  and  $k_3$  range between 1.2 and 2.0. The  $k_1$  and  $k_3$  can be determined through optimization or can be set to range within 1.2 and 2.0 in the absence of development data (Manning and Schutze, 1999).

### 2.1.4 Residual Inverse Document Frequency (RIDF)

The idea behind the RIDF is to find content words based on actual IDF and predicted IDF. The widely used methods to IDF prediction is Poisson and K mixture. However, K mixture tends to fit very well with content terms. On the other hand, Poisson deviates from the IDF remarkably and provides non-content words. Given term frequencies in triple collection, predicted IDF can be used to measure the RIDF for a triple as follows:

$$RIDF = \sum_{i \in T} \left( idf_i - \widehat{idf}_i \right) \quad (4)$$

$$= \sum_{i \in T} \left( idf_i - \log \frac{1}{1 - P(0; \lambda_i)} \right)$$

Where  $\lambda_i$  represents the average number of occurrences of term and  $P(0; \lambda_i)$  represents the Poisson prediction of  $df$  where term will not be found in a document. Therefore,  $1 - P(0; \lambda_i)$  can be interpreted as finding at least one term and can be measured using:

$$P(k; \lambda_i) = e^{-\lambda_i} \frac{\lambda_i^k}{k!} \quad (5)$$

Based on the same RIDF concept, we can moderate this to work with term distribution models that fits well with actual  $df$  such as K mixture. The definition of the K-mixture is given below.

$$P(k; \lambda_i) = (1 - \alpha) \delta_{k,0} + \frac{\alpha}{\beta + 1} \left( \frac{\beta}{\beta + 1} \right)^k \quad (6)$$

In K-mixture based RIDF we interpreted the deviation from predicated  $df$  to mark the term as a non-content term.

## 2.2 The problem as a Bag of Concepts

This section explains two BoC models which can rank triples utilizing the semantic representation of the triple collection. In particular, we employ

two widely accepted BoC models; Latent Semantic Analysis (LSA) and adoption of Log Likelihood Distance (LLD) using two corpora. The following sections describe them in detail.

### 2.2.1 Latent Semantic Analysis

This method analysed how triples in the collection can be ranked concept-wise and retrieved related to the question and answer where triples are represented in a semantic space. Such a ranking can expose the original semantic structure of the space and its dimensions (Manning and Schutze, 1999). In particular, we employed the Latent Semantic Indexing (LSI) for each collection of triples associated with the question.

### 2.2.2 Corpus based Log Likelihood Distance (LLD)

The idea behind the implementation of this method is to identify domain specific concepts (compared to the general concepts) and rank triples which contain such concepts. For this we employed a domain corpus (see Section 2.3) and a general reference corpus (see Section 2.4). The model extracts concepts which are related to the domain on the basis of their frequency in domain corpus and general reference corpus. A term that is more frequently seen in a domain corpus compared to the general reference corpus implies that the term is a concept that is used in the domain being considered (Perera and Nand, 2014b; Perera and Nand, 2014c). We utilized the log likelihood distance (He et al., 2006; Gelbukh et al., 2010) to measure the importance as mentioned below:

$$W_t = 2 \times \left( \left( f_t^{dom} \times \log \left( \frac{f_t^{dom}}{f\_exp_t^{dom}} \right) \right) + \left( f_t^{ref} \times \log \left( \frac{f_t^{ref}}{f\_exp_t^{ref}} \right) \right) \right) \quad (7)$$

where,  $f_t^{dom}$  and  $f_t^{ref}$  represent frequency of term ( $t$ ) in domain corpus and reference corpus respectively. Expected frequency of a term ( $t$ ) in domain ( $f\_exp_t^{dom}$ ) and reference corpora ( $f\_exp_t^{ref}$ ) were calculated as follows:

$$f\_exp_t^{dom} = s_{dom} \times \left( \frac{f_t^{dom} + f_t^{ref}}{s_{dom} + s_{ref}} \right) \quad (8)$$

$$f\_exp_t^{ref} = s_{ref} \times \left( \frac{f_t^{dom} + f_t^{ref}}{s_{dom} + s_{ref}} \right) \quad (9)$$

where,  $s_{dom}$  and  $s_{ref}$  represent total number of tokens in domain corpus and reference corpus respectively. Next, we can calculate the weight of a triple ( $\langle$ subject, predicate, object $\rangle$ ) by summing up the weight assigned to each term of the triple

### 2.3 Domain Corpus

The domain corpus is a collection of text related to the domain of the question being considered. However, finding a corpus which belongs to the same domain as the question is challenge on its own. To overcome this, we have utilized an unsupervised domain corpus creation based on a web snippet extraction. The input to this process is a set of extracted key phrases from a question and its answers.

### 2.4 Reference Corpus

The reference corpus is an additional resource utilized for the LLD based contextual information selection. We used the British National Corpus (BNC) as the reference corpus. The selection is influenced by the language used in the DBpedia, British English. However, what is important for the LLD calculation is a term frequency matrix. We have first performed stopword filtering on the BNC and this operation reduced the original size of BNC (100 million words) to 52.3 million words. Next, the term frequency matrix is built using a unigram analysis.

### 2.5 Triple retrieval

The model employs the Jena RDF framework for the triple retrieval. We have implemented a Java library to query and automatically download necessary RDF files from DBpedia.

### 2.6 Threshold based selection

After associating each triple with a calculated weight, we then need to limit the selection based on a particular cut-off point as the threshold ( $\theta$ ). Due

Table 1: Dataset statistics. Invalid questions are those that are already marked by dataset providers as invalid and questions where for which triples cannot be retrieved from DBpedia

	Training	Test
All questions	100	100
Invalid questions	5	10
Single entity questions	47	42
Multiple entity questions	48	48

to the absence of knowledge to measure the  $\theta$  at this stage, it is considered as a factor that needs to be tuned based on experiments. Further discussion on selecting the  $\theta$  can be found in Section 4.

### 3 Experimental framework

#### 3.1 Dataset

We used the QALD-2 training and test datasets, but removed questions which marked as “out of scope” by dataset providers and those for which DBpedia triples did not exist. Table 1 provides the statistics of the dataset, including the distribution of questions in two different question categories, single entity and multiple entity questions.

We have also built a gold triple collection for each question for the purpose of evaluation. These gold triples were selected by analysing community provided answers for the questions in our dataset. Table 1 shows the statistics for both training and testing datasets.

#### 3.2 Results and discussion

The evaluation is carried out using gold triples as described in Section 3.1. The definitions of precision (P), recall (R) and F-score (F\*) are given below:

$$P = \frac{|triples_{selected} \cap triples_{gold}|}{|triples_{selected}|} \tag{10}$$

$$R = \frac{|triples_{selected} \cap triples_{gold}|}{|triples_{gold}|} \tag{11}$$

$$F^* = \frac{2PR}{P + R} \tag{12}$$

The threshold ( $\theta$ ) (measure as a percentage from the total triple collection) value for the ranked triples was experimentally chosen to using the training

Table 2: Statistics related to the gold triple percentage in total triple collection in training dataset. The  $\mu$  represents the mean percentage of gold triples included in the total collection. The  $\sigma$  shows the standard deviation. The *Max%* and *Min%* represent maximum and minimum percentage of gold triples from the total collection respectively.

	$\mu$	$\sigma$	Max%	Min%
Single entity type	68.89	4.28	78.79	63.58
Multiple entity type	30.43	3.88	37.06	22.93

dataset. This threshold value was then used to select triples which were relevant for the testing dataset. The value was experimentally determined by using a combination of precision and recall value from the training data. For an accurate model, the precision is expected to remain constant until it starts selecting the irrelevant triples after which the precision will gradually decrease. Correspondingly, the recall value will increase until the threshold point after which the model will start selecting irrelevant triples, which will start pushing the recall value down. Hence the optimum  $\theta$  value will be the point at the maximum point for both recall and precision which is the maximum score.

Using the  $\theta$  identified from training set, we can then test the model using testing dataset. When measuring the  $\theta$  based on the training dataset it is also important to measure the proportion of gold triples compared to the number of total triples. A set of statistics related to this calculation is shown in Table 2.

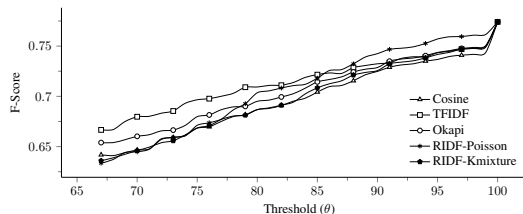


Figure 3: F-score gained by Bag of Words models plotted against threshold for questions with single entity type

According to statistics shown in Table 2 it is clear that the mean percentage of gold triples percentages

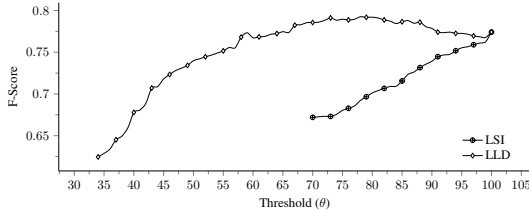


Figure 4: F-score gained by Bag of Concepts models plotted against threshold for questions with single entity type

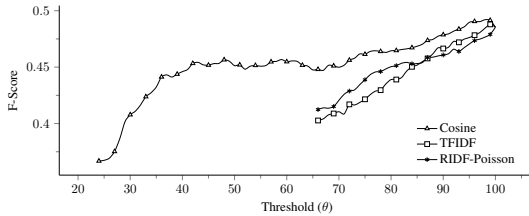


Figure 5: F-score gained by Bag of Words models plotted against threshold for questions with multiple entity types. Okapi and K-mixture based ranking methods completely failed to identify relevant triples.

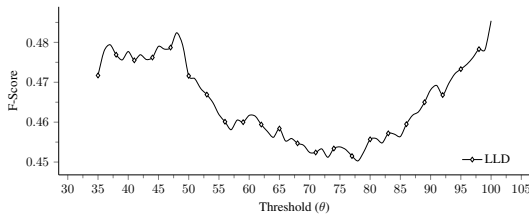


Figure 6: F-score gained by Bag of Concepts models plotted against threshold for questions with multiple entity types. The LSI method failed completely in identifying relevant triples.

are 68.89% for single entity types and 30.43% from the multiple entity types. Furthermore, the maximum and minimum percentages are also near values to the receptive mean values. This encompasses that there is a possibility to find a threshold value for both single and multiple entity types questions. Fig. 3 depicts the evaluation performed on the single entity question category in training dataset using five BoW models under investigation. The results show that the maximum F-score obtained when  $\theta$  is set to 100%. This shows that these models unable to accurately differentiate between the relevant and relevant triples. The BoW models consider only the words

Table 3: Performance of LLD on single entity type question test dataset with 78% threshold

	Precision	Recall	F-score
LLD on Test Dataset (Single Entity)	0.72	0.84	0.76

as features and therefore every entity is assigned the same importance.

The corresponding evaluation performed on the single entity question category in training dataset using BoC models is shown in Fig. 4. Latent Semantic Indexing (LSI) has performed poorly and has not managed to identify a global maximum. However, the Log Likelihood Distance (LLD) has identified a global maximum with a  $\theta$  value of 78%. Furthermore, it has also shown the expected behaviour with an increase in  $\theta$  value. When this threshold value was used to extract triples from the test dataset the results were encouraging. Table 3 shows that LLD has achieved F-score of 0.76 for the testing dataset with a 0.72 precision value. The LLD model outperforms the other models in this context mainly because it also incorporates the domain knowledge (provided through a domain corpus as explained in Section 2.2.2).

Fig. 5 and Fig. 6 depict the evaluation performed on the multiple entity question category from training dataset, for both Bag of Words and Bag of Concepts models. RIDF-Kmixture and Okapi have completely failed without any success in identifying the relevant triples. The Cosine, TF-IDF, and RIDF-Poisson have also not identified the optimum threshold (see Fig. 5). From the BoC models, the LSI method has also failed entirely. The LLD mode has identified a local maximum at  $\theta = 48$ , however the model has not behaved as expected. Furthermore, the global maximum identified at  $\theta = 100$  implies that the model can identify all relevant triples only when the total triple collection is retrieved. This confirms that although a Bag of Concepts model such as LLD performed well in the single entity type questions, none of the models performed well in contextual information selection for multiple entity type questions.

Analysis of the erroneous triples for this experi-

ment revealed that for multiple entity type questions it is important to identify the intent entity from the question. The information from the intent of the question can be used to factor in a weight correction for the triples. This leads to the study the Bag of Narrative (Cambria and White, 2014) model which is essentially based on pragmatic aspects of the language. Section 4 discusses this aspect in detail.

#### 4 Pragmatic aspect in contextual information selection

We introduce two pragmatic based concepts that need to be studied in contextual information selection approaches, derived from psycholinguistics.

- Pragmatic intent (Byram and Hu, 2013) of a question in the perspective of contextual information and,
- Pragmatic inferences (Byram and Hu, 2013; Tomlinson and Bott, 2013) that can be drawn based on already known information.

##### 4.1 Pragmatic Intent

The pragmatic intent of a question in our problem can be defined as the entity that a user is actually intending to know more about. This concept deviates from the two early approaches in query classification; Broder’s taxonomy (Broder, 2002) (classifying queries as informational, transactional, and navigational) and question typologies (Li and Roth, 2006) (determining the answer type for a question).

Consider the examples given in Table 4, where in each question multiple entities are mentioned. The entities are numbered and pragmatic intent is tagged (with code *:i*).

In  $Q_1$ , *Marc Mezvinsky* is the pragmatic intent of the question which is also the expected answer. The same rule applies for  $Q_2$ ,  $Q_3$ , and  $Q_4$ . However, in  $Q_5$  and  $Q_6$  the pragmatic intent is a part of the question ([*MI6*] and [*Natalie Portman*]), but not the answer. This variation makes it difficult to identify the pragmatic intent of a question compared to the question target identification. When presenting contextual information to the user, the information related to the pragmatic intent together with information that is shared by pragmatic intent and other entities need to be given the priority.

Table 5: Example question to illustrate the pragmatic inference used in the information elimination

$Q_7$	Which river does the Brooklyn Bridge in New York cross?
Answer	East River
Triple	(East River, flow through, New York)

##### 4.2 Pragmatic inference

The pragmatic inference is the interpreting information based on the context that it operates on. For example, a storyteller will not mention every incident or fact that happened in a narrative, thus some parts may be left for the reader to interpret using common sense knowledge, open domain knowledge, and knowledge that is already mentioned in the narrative. Applying this well-established psycholinguistic theory in our approach, we noticed several scenarios where we can improve the contextual information by eliminating information that can be pragmatically inferred and prioritizing information that needs for the context.

Consider the question ( $Q_7$ ), answer and the triple provided in Table 5. Using the question and its answer we can infer the following two facts encoded in the triples.

- $F_1$ : Brooklyn Bridge, located in, New York
- $F_2$ : Brooklyn Bridge, crosses, East River

As humans, we can infer that if Brooklyn Bridge is located in New York and if it crosses the East River, then East river must flow through New York, hence it is co-located. Therefore, the triple in Table 5 becomes unimportant for the context because it is already inferred by  $F_1$  and  $F_2$  which can be derived from question and its answer.

The pragmatic inference can also be used to prioritize the information using semantic relations that entities contain. For example consider the two scenarios illustrated in Table 6 where important contextual information can be inferred based on the semantic relationship of the pragmatic intent and entities.

In  $Q_8$  the relation between the entity “Virgin Group” and its co-founders (“Richard Branson” and “Nik Powell”) is in the form of launching a new organization. This makes the information such as cur-

Table 4: Example questions to illustrate the pragmatic intent variation in different questions. Entities are numbered and intent is marked with code *i*

#	Question	Answer
Q <sub>1</sub>	Who is the daughter of [Bill Clinton] <sub>1</sub> married to?	[Marc Mezvinsky] <sub>2</sub> <sup><i>i</i></sup>
Q <sub>2</sub>	Which river does the [Brooklyn Bridge] <sub>1</sub> in [New York] <sub>2</sub> cross?	[East river] <sub>3</sub> <sup><i>i</i></sup>
Q <sub>3</sub>	Which bridge located in [New York] <sub>1</sub> is opened on 19th March 1945?	[Brooklyn Bridge] <sub>2</sub> <sup><i>i</i></sup>
Q <sub>4</sub>	What is the highest place in [Karakoram] <sub>1</sub> ?	[K2] <sub>2</sub> <sup><i>i</i></sup>
Q <sub>5</sub>	In which [UK] <sub>1</sub> city is the headquarters of the [MI6] <sub>2</sub> <sup><i>i</i></sup> ?	[London] <sub>3</sub>
Q <sub>6</sub>	Was [Natalie Portman] <sub>1</sub> <sup><i>i</i></sup> born in the [United States] <sub>2</sub> ?	No

Table 6: Examples illustrate the use of pragmatic inference in information prioritization

#	Question	Answer
Q <sub>8</sub>	Who is the founder of [Virgin Group] <sub>1</sub> ?	[Richard Branson] <sub>2</sub> <sup><i>i</i></sup> and [Nik Powell] <sub>3</sub> <sup><i>i</i></sup>
Q <sub>9</sub>	How often was [Michael Jordan] <sub>1</sub> <sup><i>i</i></sup> divorced?	2

rent positions which are held by its co-founders to be prioritized over other information which is not strongly related to the context of the question. Next, Q<sub>9</sub> is related to Michael Jordan’s marriage. When retrieving contextual information for this question the basic information about his wives such as personal names becomes more important for the context of the question.

## 5 Related work

Benamara and Dizier (2003) present the cooperative question answering approach which generates natural language responses for given questions. In essence, a cooperative QA system moves a few steps further from ordinary question answering systems by providing an explanation of the answer. However, this research lacks the investigation to the information needs of different questions and the process of utilizing cohesive information for the explanation, without redundant text.

Bosma (2005) incorporates the summarization as a method of presenting additional information in QA systems. He coins the term, an *intensive answer* to refer to the answer generated from the system. The process of generating an *intensive answer* is

based on the summarization using rhetorical structures. Several other summarization based methods for QA such as Demner-Fushman and Lin (2006) and Yu et al. (2007) also exist with slightly varying techniques.

Vargas-Vera and Motta (Vargas-Vera and Motta, 2004) present an ontology based QA system, AQUA. Although AQUA is primarily aimed at extracting answers from a given ontology, it also contributes to answer presentation by providing an enriched answer. The AQUA system extracts ontology concepts from the entities mentioned in the question and present those concepts in aggregated natural language.

## 6 Conclusion

This study has examined the role and effectiveness of syntactic and semantic models in contextual information selection for answer presentation. The results showed that the semantic models (e.g., LLD) performed the best for single entity based questions, however the performance dropped for multiple entity questions. An analysis of the multi-entity questions showed that in order to improve performance there is a need to integrate pragmatic aspects into the ranking framework. Further work needs to be done to establish a framework to model pragmatic aspects in contextual information selection. We have already launched the development of the pragmatic framework as discussed in Section 4. Future work will introduce diverse methods of answer presentation in question answering system utilizing contextual information (Perera and Nand, 2015b; Perera et al., 2015; Perera and Nand, 2015a; Perera and Nand, 2015c).



## References

- Farah Benamara and Patrick Saint Dizier. 2003. Dynamic generation of cooperative natural language responses in webcoop. In *9th European Workshop on Natural Language Generation*, Budapest, Hungary. ACL.
- Wauter Bosma. 2005. Extending answers using discourse structure. In *Recent Advances in Natural Language Processing*, Borovets, Bulgaria. Association for Computational Linguistics.
- Andrei Broder. 2002. A Taxonomy of Web Search. *ACM SIGIR Forum*, 36(2):3–10.
- Michael Byram and Adelheid Hu. 2013. *Routledge Encyclopedia of Language Teaching and Learning*. Routledge, Taylor & Francis Group, London, UK.
- Erik Cambria and Bebo White. 2014. Jumping NLP Curves: A Review of Natural Language Processing Research. *IEEE Computational Intelligence Magazine*, 9(2):48–57, May.
- Dina Demner-Fushman and Jimmy Lin. 2006. Answer extraction, semantic clustering, and extractive summarization for clinical question answering. In *Proceedings of the 21st International Conference on Computational Linguistics and the 44th annual meeting of the ACL - ACL '06*, pages 841–848, Morristown, NJ, USA, July. Association for Computational Linguistics.
- Alexander Gelbukh, Grigori Sidorov, and Liliana Lavilla, Eduardo Chanona-Hernandez. 2010. Automatic Term Extraction Using Log-Likelihood Based Comparison with General Reference Corpus. In *Natural Language Processing and Information Systems*, pages 248–255. Springer Berlin Heidelberg.
- Tingting He, Xiaopeng Zhang, and Ye Xinghuo. 2006. An Approach to Automatically Constructing Domain Ontology. In *20th Pacific Asia Conference on Language, Information and Computation*, pages 150–157, Wuhan. Association for Computational Linguistics.
- Xin Li and Dan Roth. 2006. Learning question classifiers: the role of semantic information. *Natural Language Engineering*, 12(3):229.
- Christopher D. Manning and Hinrich Schütze. 1999. *Foundations of statistical natural language processing*. Massachusetts Institute of Technology.
- Ana Christina Mendes and Luisa Coheur. 2013. When the answer comes into question in question-answering: survey and open issues. *Natural Language Engineering*, 19(01):1–32, January.
- Rivindu Perera and Parma Nand. 2014a. Interaction history based answer formulation for question answering. In *International Conference on Knowledge Engineering and Semantic Web (KESW)*, pages 128–139.
- Rivindu Perera and Parma Nand. 2014b. Real text-cs - corpus based domain independent content selection model. In *IEEE International Conference on Tools with Artificial Intelligence (ICTAI)*, pages 599–606.
- Rivindu Perera and Parma Nand. 2014c. The role of linked data in content selection. In *Pacific Rim International Conference on Artificial Intelligence (PRICAI)*, pages 573–586.
- Rivindu Perera and Parma Nand. 2015a. Generating lexicalization patterns for linked open data. In *Second Workshop on Natural Language Processing and Linked Open Data collocated with 10th Recent Advances in Natural Language Processing (RANLP)*, pages 2–5.
- Rivindu Perera and Parma Nand. 2015b. A multi-strategy approach for lexicalizing linked open data. In *International Conference on Intelligent Text Processing and Computational Linguistics (CICLing)*, pages 348–363.
- Rivindu Perera and Parma Nand. 2015c. Realextext-asg: A model to present answers utilizing the linguistic structure of source question. In *29th Pacific Asia Conference on Language, Information and Computation*.
- Rivindu Perera, Parma Nand, and Gisela Klette. 2015. Realextext-lex: A lexicalization framework for linked open data. In *14th International Semantic Web Conference*.
- Rivindu Perera. 2012a. Ipedagogy: Question answering system based on web information clustering. In *IEEE Fourth International Conference on Technology for Education (T4E)*.
- Rivindu Perera. 2012b. *Scholar: Cognitive Computing Approach for Question Answering*. Honours thesis, University of Westminster.
- John Tomlinson and Lewis Bott. 2013. How intonation constrains pragmatic inference. In *35th Annual Conference of the Cognitive Science Society*, Berlin, Germany. Cognitive Science Society.
- M Vargas-Vera and E Motta. 2004. Aqua-ontology-based question answering system. In *Mexican International Conference on Artificial Intelligence*, Mexico City, Mexico. Springer-Verlag.
- Hong Yu, Minsuk Lee, David Kaufman, John Ely, Jerome A. Osherooff, George Hripcsak, and James Cimino. 2007. Development, implementation, and a cognitive evaluation of a definitional question answering system for physicians. *Journal of Biomedical Informatics*, 40:236–251.