Qualitative Dimensions in Question Answering: Extending the Definitional QA Task

Lucian Vlad Lita and Andrew Hazen Schlaikjer and WeiChang Hong and Eric Nyberg
{lilita,hazen+,ehn}@cs.cmu.edu, weichanh@andrew.cmu.edu
Computer Science Department, Carnegie Mellon University
5000 Forbes Avenue, Pittsburgh, PA 15213 USA

Abstract
Current question answering tasks handle definitional questions by seeking answers which are factual in nature. While factual answers are a very important component in defining entities, a wealth of qualitative data is often ignored. In this incipient work, we define qualitative dimensions (credibility, sentiment, contradictions etc.) for evaluating answers to definitional questions and we explore potential benefits to users. These qualitative dimensions are leveraged to uncover indirect and implicit answers and can help satisfy the user’s information need.

Introduction
During recent years, evaluation forums such as TREC (Voorhees 2004) have stimulated a tremendous growth of the question answering (QA) field. Successful complex architectures (Harabagiu et al. 2000) incorporate elements such as statistical components (Lita & Carbonell 2004; Ittycheriah, Franz, & Roukos 2002), knowledge resources, answer verification, planning, and theorem proving.

The main thrust in these evaluation forums has been solving factoid questions, questions that accept simple, factual answers (i.e. In what year was the first AAAI conference held?, Who was the AAAI chairperson in 1999?). Such questions require concise answers representing simple factoids: e.g. person names, dates, objects etc.

Another class of questions being explored is definitional questions. Definitional questions seek to define entities such as objects, What is ouzo?, concepts What is artificial intelligence?, and people Who is Turing? Answers to definitional questions are usually longer and more complex. For each entity there can be multiple definitions addressing different aspects of that entity. These definitions are also factual in nature and are meant to satisfy the user’s factual information needs. QA systems that can successfully answer definitional questions (Xu, Weischedel, & Licuanan 2004; Hildebrandt, Katz, & Lin 2004; Blair-Goldenshon, McKeown, & Schlaikjer 2003; Prager, Radev, & Czuba 2001) use both structured resources (e.g. WordNet, Wikipedia, Webster) and unstructured data (e.g. local corpora, the web) for fact extraction.

Due to the formulation of existing QA tasks, definitional question answering systems strive to satisfy the need for factual information. In the process of answering definitional questions, such systems filter out non-factual information, as well as marginally factual information that does not fit into a predefined view of what a definition should be.

However, it is often the case that entities (e.g. people and objects) exhibit properties that are hard to capture by standard factual methods. Moreover, there are qualitative attributes and specific factual information often associated with entities that are not captured by existing QA systems. These qualitative elements tend to complement factual data and satisfy a different kind of information need associated with definition questions.

Approach
We expand the scope of the definitional QA task by defining qualitative dimensions of answers and exploring their potential to provide users with a better understanding and more complete definitions of target entities. Answer components along these qualitative dimensions can be used to complement answers extracted using fact-based QA systems. In the following sections we explore qualitative dimensions of answers to definitional questions. These dimensions bring together known research problems, but in a new context, supporting and expanding our view the definitional QA task.

In this abstract, we explore the following dimensions as they relate to definitional questions: $D_1$ Credibility (answers from sources with varying degrees of credibility), $D_2$ Sentiment (through sentiment analysis users uncover underlying issues and problems that are inaccessible through direct factual answers), and $D_3$ Contradictions (both factual and sentiment contradictions lead to discovery of directly opposing points of view about target entities). Beyond the work presented here we investigate additional qualitative dimensions of definitional answers: $D_4$ Opinions (frequently quoted opinions about target entities), $D_5$ Relevant Topics (popular newsgroup threads and directory categories relevant to target entities), $D_6$ Temporal (frequency and validity of the answer with respect to time), and $D_7$ Geographical (specific answers may vary in frequency with geographical regions).

$D_1$ Credibility
Many question answering systems rely on the web for broad-coverage information support. Most systems do not determine the credibility of the answer source, nor do they incorporate a measure of credibility in computing the answer confidence. Credibility (Fogg et al. 2001) may also provide additional motivation for answer validation. Table 1 shows answers from a variety of sources, ranging from government agencies, univer-
ity studies, news sites, drug manufacturers and distributors, to body building sites, independent advocacy sites, and newsgroups. Knowing the relative credibility of these information sources allows users to filter out low quality information.

### D2 Sentiment

Sentiment analysis and classification (Pang & Lee 2004) identifies how sentiments are expressed in text and whether they are favorable or unfavorable towards a target topic or entity. Table 2 shows an example of actual sentiments extracted from web documents. Sentiment classification is a qualitative dimension that offers a more clear view of how entities are regarded. In definitional questions, positive and negative sentiments can co-occur in the same sentence, together with factual pieces of information (e.g. “Although vicious animals, poodles are lovely canines”). Our preliminary experiments in sentimental classification of answers to definition questions have shown a human inter-annotator classification overlap of above 75%, and a kappa statistic of above 0.45. The task consists of multi-class sentence classification into factual or sentimental (including polarity: negative or positive) classes. One of the reasons why inter-annotator agreement is good, but less than ideal is due to how we define the class “factual”. Currently it includes irrelevant facts, facts about different entities that have the same surface form as the target entity etc. In current work we focus on better defining the sentiment classification task in the context of answering definitional questions.

### D3 Contradictions

Contradictions represent another qualitative type of information that can be uncovered from a large dataset. By being exposed to frequently occurring, contradicting information about the target entity, users can uncover implicit factual information they might not have been aware of. The example in table 3 shows pairs of answers extracted from web data that are highly redundant and that would not be normally used as answers to definitional questions. Highly redundant contradicting answers give users the opportunity to uncover underlying issues which would otherwise be unidentifiable by analyzing strictly factual definitions. Contradiction in answers exposes users to new data and may reveal new investigative directions.

### Conclusions and Future Work

In this paper we present our initial work in expanding the question answering task for definitional questions. We define qualitative dimensions for evaluating answers and show how previously ignored facets in the process entity definition may help satisfy the user’s underlying information need.

Current and future work include building models for each of these qualitative dimensions and incorporating them into a fact-based question answering system. We also plan to collaborate with other research sites in order to employ existing state-of-the-art models for representing and exploiting these qualitative dimensions.

### Acknowledgements

This work was supported in part by an ARDA grant under Phase II of the AQUAINT program.

### References


