Persuasive, authorative and topical answers for complex question answering

Leif Azzopardi[1], Mark Baillie[2] and Ian Ruthven[2]

[1]University of Glasgow   [2]University of Strathclyde
Glasgow   Glasgow
G12 8QQ   G1 1XH
UK   UK
leif@dcs.gla.ac.uk   *mb,ir@cis.strath.ac.uk*

1. Introduction

The *ciqa* track investigates the role of interaction in answering complex questions: questions that relate two or more entities by some specified relationship. As in the *ciqa* 2006, our interest in *ciqa* 2007 was on contextual factors that may affect how answers are assessed. In *ciqa* 2006 we investigated factors such as topical knowledge or confidence in assessing answers through direct questioning – asking the *ciqa* assessors to directly estimate values for such variables using ordinal or categorical scales. In *ciqa* 2006 we found many useful relationships between personal contextual variables and how assessors judged answers. This year we were keen to follow this line of investigation, following a more specific research question: how do contextual variables affect the judgement of different types of information surrogate.

We created information surrogates (answers) which contained similar amounts of information but presented the information in different ways; either as neutral, topically related information (topical answers), information that was presented in such a way as to entice the read (persuasive answers), or information that was presented as coming from a named authority (authorative answers). Separately, we gathered contextual information on the assessors' preferences for such answers through the use of HTML interaction forms. Our results show differences in how assessors reacted to these different information surrogates in terms of highly they were to judge the answer as good and how likely they were to read the document containing the answer.

2. Answers

For the main *ciqa* assessment we returned a set of manually selected answers from the AQUAINT2 collection. For the interactive component of *ciqa* we presented, for each topic, a set of manually created answers to the assessor. A sample form is shown in Figure 1.

Each answer was presented within a structured presentation format: one grey field containing the answer text in red font followed by three yellow fields containing our questions to the assessor regarding the answer. These questions were to elicit (1) if the answer was a good answer (yes/no/needed to read the entire document to decide), (2) if the answer was expected – did the assessor already have an answer in mind when creating the topic and (3) what action would they perform on the answer if shown the answer by a real QA system. In *ciqa* 2006, this final question was shown to be particularly useful in determining assessors' views on the quality of an answer.

All answers contain a textual answer with ellipses to denote missing text if the answer is a fragment of a sentence e.g. "..*an appearance on Oprah or Today can shoot book sales through*

the roof…". We deliberately selected short answers, rather than whole sentences or paragraphs, to simulate the main question answering task in which short answers are preferred. Thus, what we were trying to do was investigate interaction with a good questions answering system.

Figure 1: Answer form

The content of the answers were either

1. **topical** answers. These are answers that simply attempt to answer the question by providing a neutral answer on the main topic of the question. In Figure 1, topical answers are "*In the pay-per-click model, advertisers on Google's network selected certain keywords to identify the topic of their ads.*" and "*Google expects to add 372,000 advertiser accounts over the next four years.*" which answers either the question expressed in the topic Question field or a sub-question from the topic Description field (as shown at the topic of Figure 1).

2. **persuasive** answers. These are answers that also attempt to answer the question but contain emotive words or phrases similar to those that might be used in newspapers or advertising to grab a reader's attention. In Figure 1, persuasive answers are "I*nflated amounts charged to advertisers by the search engines have prompted litigation…*" and "*Google will not divulge much information to advertisers. Google defends the practice, saying it does not want to provide*"

3. **authorative** answers. These are answers which contain a named source for the answer provided, either a person or organisation. Examples in Figure 1 are "Go*ogle will sell $6.1 billion in ads, nearly double what it sold last year, according to Anthony Noto, an analyst at Goldman Sachs.*" And "*Each [Google] advertiser pays what it's worth to them,"said Greg Stuart, CEO of the Interactive Advertising Bureau.*" Where the authorative sources are Greg Stuart and Anthony Noto and, by reference, Goldman Sachs and the Interactive Advertising Bureau.

For each topic we created 6 answers, two of each type, which were presented to the assessor. The ordering between types of answer was rotated across forms. The content of the answers was created manually: for each topic one author browsed the general Web and extracted appropriate answers so that each answer reflects, as far as possible, genuine answers. Where there were few answers on the Web we created new answers based on the answers we found.

Answers may have slightly different lengths, e.g. the authorative answers may be longer to include the source of the answers, and the types of answers may not be mutually exclusive. However, as far as possible, we tried to create good answers of each type that allowed us to compare the effect of answer type against assessor characteristics.

Assessor characteristics were gathered through HTML interaction forms, essentially questionnaires given to the assessors. Each assessor was given two forms, described separately in sections 3 and 4.

3. Preference form

Our first form elicited information on an assessor's preference for concrete versus abstract information, Figure 2. We asked a series of questions on the assessor's preference for art, literature, companions (artists, filmmakers, authors vs. scientists and politicians), or gifts.

We classified answers into reflecting two types of preference: preference for concrete, tangible items (such as non-fiction books and film, realistic art, practical gifts) and preference for abstract items with a stronger emotional appeal (such as fiction literature, genre films, abstract art and people who are artistic rather than practical). The assessor choices are given below[1] in Table 1.

---

[1] Naturally we are not making value judgments about these choices. Rather we are reflecting that different people have different tastes and preferences which may affect what information they require. As stated at the start of each form assessors could choose not to answer any of these questions.

Figure 2: Persuasive form

Our original intention was to classify assessors into assessors who may be more interested in persuasive, rather than topical, answers. However, as is seen in Table 1, nearly all assessors fell into one group – those preferring more concrete information.

Concrete	Abstract
Non-fiction books (3^2)	Fiction books (**1**)
Biopic, documentary, historical (**4**)	Comedy, horror, romance, action (**2**)
Favourite art: Photo-art, realism/impressionism (**6**)	**Favourite art:** Abstract (**0**)
Least favourite art: Abstract (**6**)	**Least favourite art:** Photo-art, realism/impressionism (**0**)
Neil Armstrong, President Kennedy, Martin Luther King (**4**)	Woody Allen, Mark Twain, Andy Warhol (**2**)
Practical gift (**6**)	Beautiful gift (**0**)
Interesting story (**4**)	Good joke (**2**)

Table 1: Assessor preferences

---

[2] We had an 'either' category which 2 assessors selected.

4. Authority form

Our second form elicited information on an assessor's preference for sources of information, Figure 3. We presented a series of 5 scenarios which we felt would be understandable and answerable by the assessors. Each scenario posed a question which required the assessor to choose which source of information they would use to answer the question.

These sources were of three types:
1. **personal experience**. These sources reflected the assessor's own personal experience. Here we did not *assume* any personal experience but suggested a method by which the assessor could gain personal experience. For example, Scenario 1 (labelled Question 1 in Figure 3) suggests that the assessor could visit the new neighbourhood; Scenario 2 suggests they could take a course of Vitamin C. In this category what is important is the assessor answers the question based on their own impressions, beliefs or experience rather than trusting or seeking another source.
2. **knowledgeable personal contact.** These sources also reflect personal experience but the personal experience belongs to named people who have more experience than the assessor. For example a trusted family physician may have more experience of vitamin supplements and knowledge upon which to base his/her recommendation. Similarly, recent graduates of a university can give more information than can be obtained in a single visit.
3. **formal information source.** The final source was information from formal sources such as Police Departments, the FDA, or a University. In this category we do not assume that the information sources are necessarily neutral.

Figure 3: Authority form

scenario	personal experience	knowledgeable personal contact	formal information source
1	2	2	2
2	1	0	4
3	2	2	2
4	1	1	3
5	1	3	2
Total	**7**	**8**	**13**

Table 2: Assessors preferences for information source

As with the preference form, section 3, we see a distinct preference – in this case for formal information sources, Table 2. This category accounts for almost half the answers given by the assessors.

5. Results

Our original intention was to classify assessors according to which might be more attracted to persuasive answers (those who prefer more abstract items) or those who might prefer more authorative answers (those who prefer formal information sources). However, as noted above the assessors did not split into groups of sufficient size to allow such an analysis.

We do present in Table 3 how the assessors judged the answers of different types and their predicted next action for each answer (accept the answer, reject and move to a better answer, or read the document). Topical answers appeared to be the best type of surrogates in the sense that assessors appeared to find it easier to judge these answers with low use of the *need more information* category and more answers being accepted without requiring further analysis. Persuasive answers, on the other hand, had higher rates of rejection (move and poor categories). Both authorative and persuasive answers had higher rates of read decision, in particular the use of a named authority in the answer seemed to lead to the assessor choosing to read a document instead of simply accepting the answer.

	good	poor	need more information to decide	accept	move	read
topical answers	**44**	10	5	**41**	6	12
persuasive answers	34	**14**	**12**	29	**10**	19
authorative answers	38	10	11	30	8	**22**

Table 3: Distribution of answer quality and predicted next action for answer types