Abstract

The recent developments in Question Answering have kept with open-domain questions and collections, sometimes argued as being more difficult than narrow domain-focused questions and corpora. The biomedical field is indeed a specialized domain; however, its scope is fairly broad, so that considering a biomedical QA task is not necessarily such a simplification over open-domain QA as represented in the recent TREC evaluations. We shall try to characterize salient aspects of biomedical QA as well as to give a short review of useful resources to address this task.

1 The task

1.1 What for whom

Question answering can be seen as an extension of Information Retrieval, and accordingly, potential users of biomedical QA are the present users of biomedical IR: the general public, medical students, health care professionals and researchers.

The general public increasingly consult knowledge resources, especially the Web, before or after seeing a doctor, for themselves or for relatives, to obtain information about the nature of a disease, the indications and contraindications of a treatment, etc. Medical students, as other students, use the Web for preparing assignments; in parallel, medical schools put an increasing proportion of their teaching material online. For health care professionals, online knowledge resources participate in continuous medical education. The traditional bibliographic databases (Medline) are now complemented with direct Web search, and discussions are seen on physicians newsgroups (and articles are published) about the best search strategies with different resources. Alper et al. (2001), looking for answers to family physicians’ clinical questions, note that “the average time to obtain an adequate answer ranged from 2.4 to 6.5 minutes.” They also mention that “One study [Ely et al., 1999] found that physicians spent less than 2 minutes on average seeking an answer to a question. Thus, most clinical questions remain unanswered.” The potential for QA technology is clear in that respect. Biomedical researchers, as other researchers, use both Web search and specialized knowledge bases (e.g., FlyBase, flybase.bio.indiana.edu).

1.2 Language specialization

Medicine is notorious for its specialized ‘jargon’. Different levels of language specialization are found depending on the sources and their intended audience. Similarly, queries use more or less technical terminology depending on their authors. The potential gap in technicity between user questions and target documents may therefore be larger than in other domains. The issue of providing the general public access to specialized knowledge through ‘consumer vocabulary’ has been given particular attention a couple of years ago with the development of general public access to the Web (see, e.g., www.nlm.nih.gov/medlineplus/). Medical QA may then be confronted with a more acute need to cater for terminological variation.1

---

1For instance, Dalmas and Rivoallan (2002) have compiled patterns for finding expressions of answers to generic questions such as “Is disease X inherited”. Many variant words are clues of an interesting passage, including (we give English equivalents) “genes”, “genetic”, “auto-
the same time, one may question the relevance of (too?) technical documents as a source of knowledge to non-specialists. The situation might be compared to that of cross-language information retrieval, where users may be confronted with documents in a language which they do not master: shall we need automated translation of the target documents into less specialized terms? As in other domains, the difficulty of finding an answer depends on the distance between the question and the available source material. Indeed, things are easier when the source material contains the answer and when its formulation is closer to the question. The amount and nature of reasoning processes involved when the formulation gap is larger still remain to be assessed.

1.3 Trustworthy sources

An extremely delicate question is that of the reliability of medical knowledge sources, especially when consulted by the general public. Whereas search engines may propose different types of Web sites as the result of a search, the kind of service provided by a QA system involves a higher degree of answer categorization and selection, and therefore may be construed as implying more endorsement of the answers. The confidence that can be assigned to a knowledge source is thus a key feature which must be taken into account. The distinct role of evidence-based medicine is also to be considered.

1.4 Availability

The volume of data and knowledge available on the Web shows great variation from one domain to another. Whereas questions related to ‘trivia’ (e.g., geography/tourism: “What is the size of the Eiffel Tower”) correspond to a well-developed area of the Web, the nature of biomedical knowledge available online must be assessed before question answering is addressed. For instance, in contrast with computer science, where technical details for many procedures are explained and discussed at length, every diagnostic or surgical technique might not be detailed on the Web. The preceding issue (reliability conditions) bears on this one, since it restricts the number of sources which are eligible for looking for answers. Another related dimension is that of language: if another language than English is considered, the quantity of online source material decreases accordingly.

In our experience (Jacquemart and Zweigenbaum, 2003) with a test set of 100 student questions in a specialized domain (oral surgery) and a language other than English (French), a thorough manual search with Google was unable to obtain relevant answering documents within the top five hits for 40% of the questions. This contrasts with the 75–95% of answers obtained by (Alper et al., 2001) on a set of 20 clinical questions when searching ‘electronic medical databases’.

2 The resources

Medical information processing has a long tradition of compiling large-scale resources for dealing with medical knowledge and information. We review a few of these resources which are relevant to biomedical question answering.

2.1 Health information sources

Web directories have been developed to help users find classified information. In a similar way, health information gateways now propose to perform this task for the biomedical domain. For instance, the CISMeF quality-controlled health gateway (www.chru-rouen.fr/cismef/, Darmoni et al. (2000)) selects resources according to quality criteria (e.g., www.medcertain.org), which addresses the above-mentioned trust issue; it indexes medical Web sites with a structured thesaurus (the MeSH thesaurus, NLM (2001)), which helps to identify potentially relevant documents more precisely.

These directories are undoubtedly a place to visit if one is looking for precise, endorsed online biomedical information. Among the types of resources which can be found, teaching material occupies a good rank. An example is the project which has been lead for a couple of years by a consortium of French medical universities to create an online French-language Virtual Medical University (UMVF, Le Beux et al. (2002)). Another type of resource, practice guidelines, compile the most
up-to-date and scientifically verified (‘evidence-based’) clinical knowledge.

A wealth of biomedical information also exists on commercial sites and CDROMs: drug knowledge bases (e.g., in France, the Vidal drug monographs), encyclopedias and other resources (e.g., www.dynamicmedical.com) provide authoritative knowledge which is precious as a source of answers to questions. Their online versions however often have restricted access.

2.2 Types of questions

Ely et al. (2000) have studied 1396 questions collected from more than 150 physicians, mainly family doctors, and propose a taxonomy of generic clinical questions. The main question types are listed in table 1. Such a taxonomy, according to the British Medical Journal’s comments, “has four potential uses: to organise large numbers of real questions, to route questions to appropriate knowledge resources by using automated interfaces, to characterise and help remedy areas where current resources fail to address specific question types, and to set priorities for research by identifying question types for which answers do not exist.”

| What is the drug of choice for condition X? |
| What is the cause of symptom X? |
| What test is indicated in situation X? |

Table 1: Most frequent generic questions derived from questions by primary care doctors (from Ely et al., 2000).

2.3 Linguistic and terminological resources

Open-domain QA draws liberally on linguistic resources: lexicons indeed, but also the ubiquitous WordNet thesaurus. The biomedical domain, at least in the English language, proposes its own specialized lexicon as well as plenty of structured thesauri.

The resource here is the Unified Medical Language System (UMLS. Lindberg et al. (1993)). Its main component is the Metathesaurus, which compiles and cross-references one hundred biomedical terminologies (in version 2003AA: more than 800,000 concepts and 2,000,000 strings), with their hierarchical and transversal relations. Its Semantic Network adds a common structure above these imported terminologies. Additionally, its Specialist Lexicon provides a large English lexicon with an emphasis on biomedical words, including derivational knowledge. Tools have been built around the UMLS to address terminological variation (e.g., MetaMap, Aronson (2001)). The UMLS and its companion tools can be obtained free of charge from the US National Library of Medicine. While the Specialist Lexicon and most of the terminologies included in the Metathesaurus are in English, many of these terminologies are international and exist in other languages. Medical lexicons are also being created in other languages (Weske-Heck et al., 2002; Zweigenbaum et al., 2003).

The UMLS can, to some extent, be compared with WordNet: it provides terms, synonyms, hierarchical relations, but most of its organization and structure is that of a thesaurus rather than that of a formal ontology. Galen is an actual medical ontology, expressed in a description logic formalism (Rector et al., 1997). It precisely defines several thousand medical concepts based on more primitive concepts and relations. Produced by the Galen European project, it can be downloaded from the OpenGalen web site (www.opengalen.org). Galen does propose sound, formal concept descriptions; its use is also more complex than that of a simple term hierarchy.

The Gene Ontology is a more recent controlled vocabulary dedicated to genomics, more specifically the description of gene products and their associated molecular functions, biological processes and cellular components. The Gene Ontology can be downloaded from www.geneontology.org. It provides terms, synonyms, multiple hierarchies with explicit “is-a” and “part-of” relations.
3 Conclusion

The biomedical domain raises new challenges for question-answering systems, but at the same time already proposes some resources to address these challenges: quality-controlled health information gateways offer a thorough indexing of trustworthy biomedical sources (section 2.1); taxonomies of question types rank and categorize interesting questions, taking into account their frequency of occurrence (section 2.2); biomedical lexicons, terminologies and ontologies are there to help manage domain-specific terms and concepts (section 2.3).

Although a few questions in the past TREC QA tracks have touched on the medical domain, no specific evaluation of medical QA has yet been performed. This is to happen in the French QA evaluation initiative EQueR (Grau, 2002), which is to take place in the coming year, where a medical QA track is planned. The preparation and execution of this track will tell us more about how the above considerations materialize in actual systems.

References


