From: AAAI Technical Report SS-03-07. Compilation copyright © 2003, AAAI (www.aaai.org). All rights reserved.

A Technique to Resolve Contradictory Answers

Catherine Blake

School of Information and Computer Science
University of California, Irvine
cblake@ics.uci.edu

Abstract

For users in information intensive environments, it is becoming less likely that a single document will provide an answer to a question. Our goal is to develop techniques that unify redundant and contradictory answers, which are inevitable when multiple documents contain an answer to a question. When users make consequential decisions, such as users in medicine and public health, it is important that they play an active role in the process, and that the answer reflects all of the available evidence. Our study of scientists as they answered research questions based on existing biomedical literature motivated the hierarchical question answering approach that we outline in this position paper. The proposed approach, which we call Information Synthesis (IS), is suitable for the class of questions that have a quantitative answer. We are currently developing a system that supports our proposed solution.

Introduction

The amount of information available electronically continues to increase at an alarming rate. The motivation behind both information retrieval (IR) and question answering (QA) systems is to support interactions with users throughout their information-seeking activities. In contrast to traditional information retrieval systems that retrieve a set of documents that satisfy an information need, a question-answering system retrieves a specific section of text within a document that answers a user's question. In both IR and QA systems, designers assume that the information that is required to address a user's information need is explicit in at least one document within the corpus. This position paper explores challenges that arise when a corpus contains multiple answers to single question.

Users operating in information intensive environments often face the challenge of integrating information from multiple sources, such as when news articles report different facts. In this case, both temporal factors and the opinion of the authors who provide the information can influence the quality of the candidate answers. We can also attribute differences in scientific findings to these factors. However, we propose that the complex nature of the scientific phenomenon described in an article explains much of the variance. Regardless of the source, the importance of decisions made in scientific domains, require that users resolve the differences reported in studies. The answer to a question should therefore consider all of the available evidence related to a user's initial information need, in this case expressed as a question. Due to the critical nature of decisions made in medicine and public health, the user is an integral component of our approach.

Consider the question *What is the relationship between alcohol consumption and breast cancer risk?* A recent analysis that explored this risk-factor and disease relationship identified more than seventy articles that each proposed an answer to the alcohol-breast cancer question (Ellison et al. 2001). To unify these differences Ellison and his colleagues used a set of statistical techniques called meta-analysis. Originally developed for the agricultural community, these techniques have undergone a resurgence in medical and public health domains due to the increased popularity of evidence-based medicine. A meta-analysis is a type of literature review that provides a quantitative summary of empirical evidence reported in multiple independent studies (Lipsey and Wilson, 2000), and when based on a set of randomized clinical trials provides the strongest form of medical evidence, (Harbour and Miller, 2001; Hunter and Schmidt, 1990). Our approach focuses on semi-automating the quantitative components of the meta-analysis process.

An Information Synthesis Scenario

Consider again the question *What is the relationship between alcohol consumption breast cancer risk?* We indicate this original, high-level question in Figure 1, Level 4. Based on first principals of epidemiology theory we can decompose this Level 4 question into two distinct Level 3 questions. We focus on *Do people with breast cancer consume significantly more alcohol than people in a similar population who do not have breast cancer?* A significantly higher rate of alcohol consumption rate by people with breast cancer, than people without breast cancer suggests a co-occurrence between the user behavior and breast cancer and that alcohol should be further explored as risk factor for breast cancer.

A traditional meta-analysis would consider only those studies that initially set out to explore the risk factor and disease relationship. However, scientists have established

Copyright © 2002, American Association for Artificial Intelligence (www.aaai.org). All rights reserved.

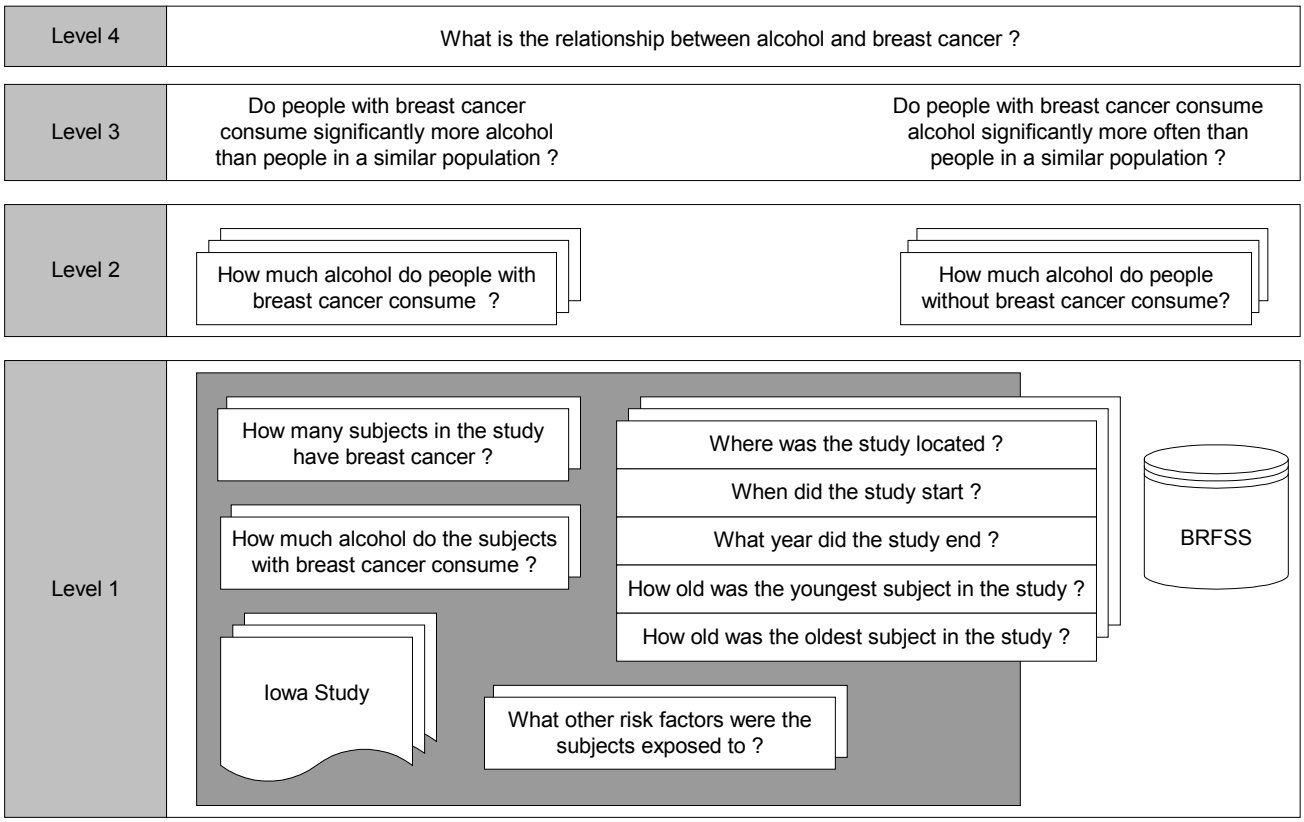

Figure 1 – **Decomposition of the original question into a hierarchy of more specific questions.** Information from full-text articles and an external database (the Behavioral Risk Factors Surveillance System) provide answers to questions at lower levels of the hierarchy. Scientists can combine lower level answers to address higher-level questions. Augmenting facts in scientific literature with the external database enables scientists to consider currently unused information from text.

that journal editors are more likely to publish an epidemiological study that identifies a relationship between a risk factor and a disease, than a study that does not identify such a relationship (Easterbrook et al. 1991). A meta-analysis based on this systematically biased sample of studies could therefore overestimate the true impact of the candidate risk factor (Begg and Berlin, 1988). We base our approach on the premise that studies that report risk-factor exposure rates as secondary information in an article are subject to different types of publication bias than studies that report the finding as primary information. Our definition of secondary information is facts that are not in the title, abstract, or keywords of an article. To accurately answer our initial question, we therefore propose a non-traditional meta-analysis, which considers studies that are not primarily about the risk factor and disease. In our scenario, this manifests as breast cancer studies that report alcohol consumption, even though their primary topic does not relate to alcohol and breast cancer.

We can now decompose the first question in Figure 1, Level 3, into the two questions, *How much alcohol do people with breast cancer consume* and *How much alcohol do people without breast cancer consume*. In a traditional meta-analysis, the answer to each of the Level 2 questions would be within a single document. We could then use the differences between these answers to address questions at higher levels of the hierarchy. Our secondary goal is to incorporate currently used information from studies that did not set out to explore the risk factor and disease relationship. To achieve this goal, we augment the information reported in each study with information from an external database, because studies that do not report the risk factor and disease relationship the primary topic generally lack a control group estimate, which is required for the analysis. In this scenario, we estimate the alcohol consumption of a similar population using the age, gender, time, and location- specific alcohol consumption rate within the Behavioral Risk Factors Surveillance (BRFSS) system (BRFSS, 2002). The BRFSS provides the results of an annual survey that explores health behaviors of US residents, such as exercise habits, alcohol and tobacco consumption.

An individual study or the external database provides the answers to the lowest level questions in our decomposition (Level 1). In Level 1, we make similar assumptions to existing QA systems, that each document has only a single answer to a question. The user would

play an integral role during this process, such as to verify lower level answers.

After recursively answering each of the questions in the lower levels, the system calculates the difference between the rate of alcohol consumption reported in breast cancer studies, and the rates reported in the BRFSS comparison group (after controlling for study size). A significant difference would provide strong evidence that alcohol consumption increases breast cancer risk, thus answering the users initial question.

Discussion

The multi-level query reformulation that we describe provides a conceptual framework to resolve contradictory answers for the next generation of question answering systems. Using a manual version of this approach, others have explored research questions regarding the correlation between smoking and impotence (Tengs and Osgood, 2001). There are clearly opportunities to automate components of this approach, such as using existing question-answering technology to identify lower level answers, and encoding procedural knowledge of meta-analysis to synthesize these answers. In contrast to a similar deductive question answering approach demonstrated in the geographic domain, our approach requires that a user make decisions regarding the validity at each level, rather employing a theorem proving module (Waldinger et al, 2002). For consequential decisions, such as those in public health and medicine, we maintain that a user should verify answers to questions at lower levels. In contrast to their logic-based approach, our approach is suited to questions that require a quantitative answer.

Information Synthesis provides an opportunity to re-use answers found at each of the different hierarchical levels. This re-use is supported by our study of scientists as they used biomedical literature to answer research questions, where we found that despite very different review topics, similar information items were used (Blake and Pratt, 2002). For example, answers to questions regarding the population group (including age, ethnicity and gender of the subjects) and the study (such as the design, and number of participants) were required for a review that explored the reliability of spinal palpatory procedures, a meta-analysis that explored the relationship between smoking and impotence and a study of hypertension treatments. This regularity suggests that users can incorporate answers within the decomposition hierarchy to answer different higher-level questions.

We have completed the user study that motivated this approach (Blake and Pratt, 2002). Promising results from our prototype system (Blake, Pratt, and Tengs, 2002) have spurred our enthusiasm towards developing a full implementation to support the scenario described in the previous section. Our current implementation automatically identifies answers to a set of pre-defined questions from full-text journal articles, which are required to conduct a meta-analysis on the relationship between smoking and breast cancer. After manual verification, the system then automatically conducts the quantitative component of a meta-analysis to synthesize these lower level answers.

Meta-analytic techniques are not constrained to medical questions. Figure 2 provides a decomposition to a question from the business domain.

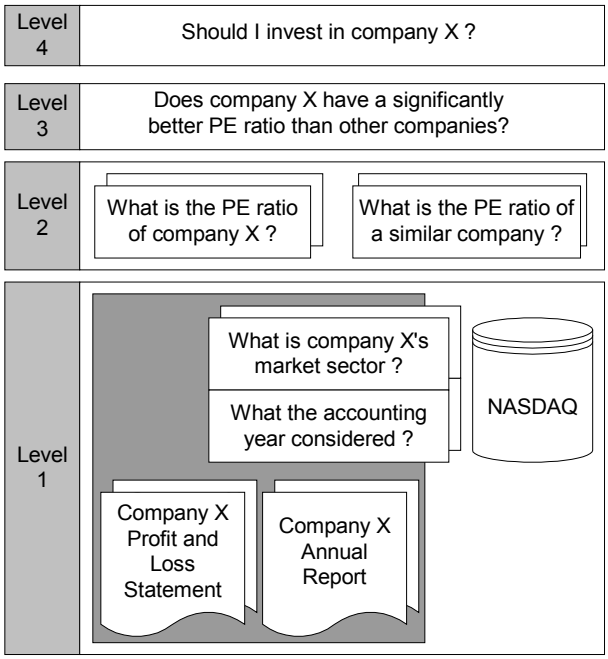

Figure 2 – Decomposition of the question *Should I invest in company X?*

We are not the first to recognize that a corpus can contain multiple answers to a question. The AskMSR system determines which answer to provide to a user, based on the frequency of each proposed answer (Banko et al. 2002). Meta-analysis combines each study finding based on the inverse of the study variance, rather than by the most frequently occurring answer, as suggested in AskMSR.

Other work has considered multi-perspectives with respect to news information (Pustejovsky, Wiebe, and Maybury, 2002). In contrast to identifying subjective aspects of an article, our work focuses on integrating quantitative information reported from multiple sources. We also do not address the complex task of natural language generation to provide a summary, such as proposed in PERSIVAL to generate patient specific summaries (McKeown et al. 2001).

Conclusions

We have proposed a new quantitative approach to address the challenges that arise when a corpus contains redundant and contradictory answers to a question, which is inevitable when multiple documents contain an answer to a user's question. Our hierarchical decomposition of a

question enables re-use of answers at lower levels to answer a variety of questions at higher levels. Our approach supports the active role that users must play when making consequential decisions, such as decisions in medicine and public health and enables scientist to incorporate information that existing meta-analytic techniques do not consider.

Acknowledgements

A dissertation award #8GB-0175 from the California Breast Cancer Research Program funded this work. The author would like to thank Drs Wanda Pratt, Catherine Carpenter, and Tammy Tengs for earlier discussions of this approach.

References

Banko, M.; Brill, E.; Dumais, S.; and Lin, J. 2002. AskMSR: Question Answering Using the Worldwide Web. In S. M. Harabagiu, and V. Chaudhri (Eds.), *Mining Answers from Texts and Knowledge Bases, AAAI Spring Symposium, Technical Report SS-02-06*. Stanford: AAAI Press.

Begg, C.; and Berlin, J. 1988. Publication bias: a problem in interpreting medical data. *Journal of the Royal Statistical Society*(151): 419-463.

Blake, C.; and Pratt, W. 2002. Collaborative Information Synthesis. In *Proceedings of American Society of Information Science and Technology,* 44-57. Philadelphia, PA.

Blake, C.; Pratt, W.; and Tengs, T. 2002. Automated Information Extraction and Analysis for Information Synthesis. In *Proceedings of American Medical Informatics Association Fall Symposium*. 976. San Antonio, TX.

Easterbrook, P. J.; Berlin, J. A.; Gopalan, R.; and Matthews, D. R. 1991. Publication bias in research. *Lancet*, 337: 867-872.

Ellison, R. C.; Zhang, Y.; McLennan, C. E.; and Rothman, K. J. 2001. Exploring the Relation of Alcohol Consumption to Risk of Breast Cancer. *American Journal of Epidemiology*, 154(8): 740-747.

Harbour, R.; and Miller, J. 2001. A new system for grading recommendations in evidence based guidelines. *BMJ*, 323: 334-336.

Hunter, J. E.; and Schmidt, F. L. 1990. *Methods of Correcting Error and Bias in Research Findings*. Newbury Park, CA: Sage.

Ingelfinger, J. A.; Mosteller, F.; Thibodeau, L. A.; and Ware, J. H. 1994. *Biostatistics in clinical medicine* (3rd ed.): McGraw-Hill Inc.

Lipsey, M. W.; and Wilson, D. B. 2000. *Practical Meta-Analysis*: Sage.

McKeown, K.; Chang, S. F.; Cimino, J.; Feiner, S.; Friedman, C.; Gravano, L.; Hatzivassiloglou, V.; Johnson, S.; Jordan, D.; Klavans, J.; Kushniruk, A.; Patel, V.; and Teufel, S. 2001. PERSIVAL, a System for Personalized Search and Summarization over Multimedia Healtcare Information. In *Proceedings of 1st Joint Conference on Digital Libraries,* 331-340. Roanoke, VI.

Pustejovsky, J.; Wiebe, J.; and Maybury, M. 2002. Multi-perspective and Temporal Question Answering. In *Proceedings of Third International Conference on Language Resources and Evaluation (LREC 2002), Workshop on Question Answering: Strategy and Resources*. Canary Islands, Spain.

Tengs, T.; and Osgood, N. D. 2001. The link between smoking and Impotence : Two Decades of Evidence. *Preventive Medicine*, 32(6): 447-452.

Waldinger, R.; Reddy, M.; Culy, C.; Hobbs, J.; and Dungan, J. 2002. Deductive Response to Geographic Queries. In *Proceedings of GIScience*. Boulder, CO.