Finding Nutrition Information on the Web: Coverage vs. Authority

Susan G. Doran

Department of Computer Science and Engineering, University of South Carolina, Columbia, SC 29208.Sue_doran@yahoo.com

Samuel de Ycaza

ClaflinUniversity, Orangeburg, SC.samueldeycaza@yahoo.com

Caroline M. Eastman, Corresponding Author

Department of Computer Science and Engineering, University of South Carolina, Columbia, SC 29208.

eastman@engr.sc.edu

Bernard J. Jansen

School of Information Sciences and Technology, The Pennsylvania State University, University Park, PA, 16801.jjansen@acm.org

The ability of different types of web sites to provide relevant information in response to queries in the area of nutrition is investigated here. Nutrition related queries from users of the Excite search engine were posed to a number of nutrition Web sites in order to assess whether or not topical matches were found. The queries were also submitted to four general-purpose search engines. Surrogate users were used to assess the actual relevance of topical matches. Matches were retrieved for about 36% of the queries on the nutrition sites; the range for different sites went from a low of 8.6% to a high of 68.6%. For the general search engines, matches were retrieved for about 92% of the queries.

Almost all of the topical matches in both cases were judged relevant.

Introduction

Health information is one of the major categories of information that web users look for. We are all concerned with health, including both the prevention and management of health problems. We focus here on a major subcategory of health-related information, nutrition. How easily users searching for nutrition information on the Web can find the information they desire is investigated here. This analysis may provide a better understanding of the nutrition information needs of Web users and the design of useful nutritional knowledge bases.

The growth of the Web has led to increased interest in end user information retrieval (IR) systems, as manifested by the proliferation of Web search engines. There has also been a growth of indexing and classification systems and widespread public use of Web IR systems (Jansen & Pooch, 2001). This growth has also led to an expanding variety of search topics including locating information on nutrition. IR is a non-trivial problem; what is relevant is decided by the user from session to session, may change from time to time, and is heavily dependent on individual judgments (Saracevic, 1975). Making judgments of information quality and authority is difficult for most users because overall, there is no quality control mechanism (Reih, 2002). Judging quality is especially difficult in the domain of nutrition, where conflicting claims are abundant.

In the following sections, we state our research question, address related work, describe the methodology of our study, and provide an overview of the results. This is followed by a discussion and the implications for user information seeking. The work presented here is an extension of work previously presented by Doran et al. (2003) and Doran (2002).

Research Question

Our research objective was to determine to what extent specialized nutritional sites are providing the information that general web users are seeking. Many of these sites, but not all, are carefully constructed and regulated to provide authoritative information on nutrition. Our question specifically addresses the question of coverage within the field of nutrition.

Related Work

Work related to that presented here includes studies of medical sites in general, nutrition sites in particular, analysis of user queries in these areas, and methods to assess the quality of information retrieved. Spink et al. (2004) provide a general overview in the context of their analysis of queries to general search engines. They compared web searches submitted to several general engines over a period of years. For example, the percentage of queries to Excite on medical topics declined from 9.5% in 1997 to 7.5% in 2001. (They attribute this to the increase in commercial use.) Many of the queries in the general area of health were about weight issues; presumably some of the other categories included some nutrition-related queries as well. So this study documents the use of general search engines to search for nutrition-related information. There have been some studies examining the use of specific nutrition sites; examples include McCray et al. (1998) and Barnas & Kahn (1999). The Pew Internet Project has examined the use of internet resources for medical information (Pew Internet & Family Life, 2002, 2003). They found that 65% of health information seekers were looking for information on exercise, nutrition, or weight control. (The categories they used were not mutually exclusive, so searchers could be in more than one category.)

At least one other study has compared health and nutrition sites with general search engines (Atlas, 2001); this study found that medical students had more success searching in general search engines than more specialized sites. This finding is consistent with that reported here. New tools for search web medical resources were the topic of research in which the lack of advanced search capabilities and the limitations in precision, number, and export options of the results supplied were described (Aguillo, 2000). There has been substantial work involving the quality of health and nutrition information on the web, both to evaluate existing information and to develop procedures for providing authoritative and correct information (McLeod, 1998; Eysenbach et al., 2000; Fallis & Fricke, 2002).

Research Methodology

This study utilized 70 nutrition queries submitted to 13 nutrition sites, 1 nutrition metasite, and 4 general search engines. The search results were evaluated for relevance. The matches and relevance measures for the two methods of information seeking, specialized niche sites and general search engines, were then compared. The overall ability of the sites linked to by the metasite (Tufts Nutrition Navigator) was then investigated.

Query Selection

The 70 queries used (see Table 1) are from an Excite transaction log containing actual user requests. An original list of 100 terms selected automatically was screened by a registered dietitian. A college nutrition textbook, Zeman's Clinical Nutrition and Dietetics, was examined to confirm the relationship between the terms comprising the queries and the field of nutrition (Zeman, 1983). In the field of nutrition, Zeman's text is considered a reliable standard of reference.

Query	Frequency	Query	Frequency
recipes	247	"food posters"	18
weight loss	105	catering services food	18
FAT BURNING FOODS DIET	84	Will I lose weight eating 2090 cals per day?	17
Diabetes	63	greek food	17
weight watchers	57	weightwatchers	17
Fitness	52	easter recipes	16
breast cancer	41	health and nutrition MLM'S	16
herbs	39	nutritional supplements	16
Cancer	36	quick recipes	16
nutrition	36	american cancer society	16
fruits of vietnam	33	prostate cancer	16
Food	30	passover recipes	16
Diabetes, AND Israel	29	Cancer Research	16
indian food calorie chart	26	Prader willi kids	16
fat burning foods	26	weight training	15
articles on children nutrition	24	weightloss	15
ovarian cancer	23	eating disorders	15
Nutrition Games	22	long island ice tea recipes	15
Foodtv.com	22	Fruit	15
health food birmingham alabama	22	crop and food	15
boneless chicken breast recipes	22	"recipes using lavender"	15

ovarian cancer	23	eating disorders	15
Nutrition Games	22	long island ice tea recipes	15
Foodtv.com	22	Fruit	15
health food birmingham alabama	22	crop and food	15
boneless chicken breast recipes	22	"recipes using lavender"	15
truth in nutrition ads	22	Paulines health food store	14
lung cancer	22	www.foodtv.com	14
Indian dishes and their nutritional value	22	recipes for beef and macaroni	14
+weight +training +nutrition	22	low fat recipes	14
breastfeeding	22	herbal remedies	14
potassium	21	"monosodium glutamate"	13
exercise	21	foodtv	13
vitamins	20	'physcological disorders'	13
chinese recipes	20	cookie recipes	13
food guide pyramid	20	salt	12
sweet recipes	20	www.diabetes.com	12
"What Are Carbohydrates."	18	recipes and healthy	12
protein	18	chinese recipes	12
cholesterol	18	meat nutritional cotent	12

Web Site Selection

Sites (see Table 2) were selected from Tufts University Nutrition Navigator (Tufts). Tufts is an online rating and review guide designed to assist users in sorting through nutrition information on the Web and finding accurate, useful nutrition information (Tufts, 2005). Tufts uses a twenty five-point scale to rate the sites: twenty points for content and five for usability. A stratified sample of sites was chosen with some sites having high ratings and others low ratings. Tufts itself was searched to determine how many of the web sites it includes contained information using the query terms; this information complements that obtained by examining individual web sites. In addition, the nutrition queries were posed to four popular general search engines: Excite, Google, MSN, and Teoma. The site searches were performed in 2002. The extension to general search engines and the metasite searches was performed in 2003. We comment on the possible impact of recent changes in the final section of this paper.

Study Design

We discuss the two parts of this study individually. The first compared the performance of individual sites and search engines in order to compare their relative effectiveness in retrieving nutrition information as represented by the sample of queries from the Excite query log. The second considered the overall coverage of nutrition sites in general and addressed the question of how widely topics are covered across specialized sites.

Each query was used for a search in each nutrition site and search engine. Topical relevance was evaluated by one or more of the authors. Three authoritative users with expertise in nutrition acted as surrogate users to review the resulting matches. A match was classified as relevant if all three users agreed that it was responsive to the query. Results were not evaluated for quality or authoritativeness. The three users were in agreement 100% of the time. We attribute this high degree of agreement to the relatively focused task and high degree of expertise. Overall results are in Table 2.

Website/Search Engine Name, Address	Tufts Rating	Queries with Matches	Percent Matches	# Relevant Matches	% Relevant Matches
Google, http://www.google.com	N/A	67	95.7%	67	100%
MSN Search, http://www.msn.com	N/A	67	95.7%	67	100%
Teoma, http://www.teoma.com	N/A	66	94.3%	66	100%
Excite, http://www.Excite.com	N/A	62	88.6%	62	100%
Food and Drug Administration, http://www.fda.gov/default.htm	24	48	68.6%	43	89.6%
iVillage–Diet & Fitness, http://www.ivillage.com/diet	17	42	60%	42	100%
American Cancer Society, http://www.cancer.org	23	38	54.3%	33	86.8%
American Dietetic Association, www.eatright.org	22	29	41.4%	29	100%

http://www.ivillage.com/diet					
American Cancer Society, http://www.cancer.org	23	38	54.3%	33	86.8%
American Dietetic Association, www.eatright.org	22	29	41.4%	29	100%
Atkins Nutritionals, http://atkinscenter.com/dev/	12	25	35.7%	25	100%
Health World Online, http://www.healthy.net	12	25	35.7%	25	100%
Food, http://lancaster.unl.edu/food	22	24	34.3%	24	100%
Consumer Information Center, http://www.pueblo.gsa.gov	24	24	34.3%	22	91.7%
Center for Disease Control and Prevention, http://www.cdc.gov	21	23	32.9%	23	100%
Weight Focus, http://www.weightfocus.com	17	23	32.9%	23	100%
Prevention, http://www.prevention.com	18	20	28.6%	20	100%
The Diet Doctor, http://www.thedietdoctor.com	13	9	12.9%	9	100%
Nutrition Resource.Com, www.nutritionresource.com	17	6	8.6%	6	100%
Average: Search Engines Only	N/A	64.5	93.6%	64.5	100%
Average: Nutrition Sites Only	18.6	25.84	35.5%	24.9	90.9%

Results

In the second phase, each of the 70 queries was entered into the Tufts Nutrition Navigator to determine the number of matching sites which contained at least some information matching the given query. With the exception of the range of 151 to 500 sites, in

which no queries showed up, the results are much as might be expected. Many of the query terms were not indexed in any of the sites. A few queries were covered by most sites. The results are shown in Table 3.

Number of Queries	Sites Retrieved
9	>501
0	151 – 500
7	51 – 150
10	11 – 50
20	1 – 10
24	0

Discussion

Detailed analysis of query performance by sites is given in Doran (2002). The overall performance range of nutrition sites was from a low of six queries retrieving matching results from Nutrition Resource.Com to a high of 48 matches from the FDA web site. All of the four general search engines retrieved a higher number of topical matches than the nutrition sites; Google and MSN search retrieved matches to 67 queries, Teoma retrieved matches to 66 queries, and Excite retrieved matches to 62 queries.

Several specific problems were observed with the sites. For example, the lack of an effective search engine prohibited iVillage from retrieving a match for *weightwatchers*, yet there was information on the site about "weight watchers" spelled as two words. The Diet

Doctor did not have a search option at all. Health World Online's search option was frequently "unavailable". The search for *weight loss* resulted in no matches on the Weight Focus site, but several sources of information on weight loss were found on this site through exploration of links.

Another performance problem observed with some nutrition sites was the inability to handle "incorrect" syntax. For example, the eight queries containing quotes, commas, '+' or 'AND' caused syntax errors on the Centers for Disease Control site. All queries containing quotes, periods and the '+' operator resulted in syntax errors on Nutrition Resource.Com.

The lack of basic nutritional information was another problem observed. For example, diet prevents many forms of cancer and adequate nutrition is critical to those undergoing treatment for cancer, yet the American Cancer Society retrieved matches to 38 queries, of which 33 contained relevant items. On Health World Online, the query "salt" was unmatched, yet it would be reasonable to expect a match to this on a nutritional site. Atkins Nutritionals did not retrieve matches for basics such as *food guide pyramid*, *potassium*, or most of the cancer queries. On Prevention, *food guide pyramid* was found but not a definition or a picture. There were no matches to the query *recipe* by Weight Focus. On the American Dietetic Association site was no match for *potassium*, which one would expect to see on this site. If there is information on the ADA site about potassium, there is no indication as to which, if any, links contains it. The lack of basic nutritional information was observed with many sites.

From the metasite searches in Tufts, it is apparent that the failure of some nutrition sites to locate information for some of the queries is not simply a result of the selection of sites. None of the sites included at that time in Tufts would have contained responses matching those queries. Examination of the terms matching large numbers of nutrition sites shows that these are fairly general nutrition terms, e.g., *nutrition, food, recipes*, and *protein.* The terms with no matching sites are generally (but not always) phrases and often fairly specialized, e.g., *boneless chicken recipes, health and nutrition MLMs*, and *Indian food calorie chart.* Three contain misspellings, e.g., *meat nutrition cotent.* The queries in the middle ranges are generally single terms or two-word phrases; they are often requesting information that would be expected in a nutrition web site, e.g., *salt, potassium,* and *cholesterol.*

Conclusions

The average percentage of queries answered by the nutrition sites and search engines was 35.5% and 93.6%, respectively. If the

Excite queries compare to those typically made to these sites, then improvements to the search engines and contents of the nutritional sites are needed. Based on the high percentage of query matches supplied by the general search engines, users have a better chance of locating relevant information using general search engines rather than using niche sites. However, these search engines provide no evaluation of the quality of information contained within the sites retrieved even when they use link or other information to rank apparently more authoritative sites more highly. Current nutrition information seekers thus have a choice between small, controlled niche sites with limited information and large uncontrolled search engines with vast amounts of information.

There are several limitations of the current work which should be addressed. The searches were performed a couple of years ago. More recent, but not complete, searches indicate that the general pattern has not changed, however. Future work will include further analysis of the existing data and its comparison to more recent data. Some of the issues that need to be addressed include the relationship of site focus, if not nutrition in general, to the results, the relationship of the Tufts site ratings to the retrieval results, and consideration of other general search engines. Future research will also focus on possible methodologies to link the nutritional information searchers desire with the organization of nutritional knowledge collections. Unfortunately, the Tufts Nutrition Navigator is no longer operational, so other resources will be used for further research.

References

Aguillo, I. (2000). A new generation of tools for search, recovery and quality evaluation of World Wide Web medical resources. *Online Information Review, 2*. Retrieved February 15, 2002, from Emerald database http://pippo.emeraldinsight.com

Atlas, M. C. (2001) First-year students' impressions of the Internet. *Medical Reference Services Quarterly, 20*, 11-25.

Barnas, G. P. & Kahn, C. E. (1999) Assessing consumer's interest in Internet-based health information. *AMIA'99*. American Medical Informatics Association.

Doran, S. G. (2002). *Nutrition information retrieval by web users*. Masters thesis. Department of Computer Science and Engineering, University of South Carolina, Columbia, South Carolina.

Doran, S. G., Eastman, C. M., & Jansen, B. J. (2003) Nutritional information on the Web: An analysis of information sought and

information provided. *Proceedings of the 2003 IRMA International Conference, Philadelphia, Pennsylvania, May 18-21, 2003*, (pp. 106-108).

Eysenbach, G. & Kohler, C. (2002) How do consumers search for and appraise health information on the World Wide Web? Qualitative study using focus groups, usability tests, and in-depth interviews. *BMJ 24*, 573-7.

Fallis, D. & Fricke. M. (2002) Indicators of accuracy of consumer health information on the Internet: a study of indicators relating to information for managing fever in children in the home. *Journal of the American Medical Informatics Association 9*, 73-9.

Jansen, B. J., & Pooch, U. (2001). Web user studies: a review and framework for future work. *Journal of the American Society for Information science and Technology, 52*(3), 235-246.

McCray, A. T., Loane, R. F., Browne, A. C. & Bangalore, A. (1998) Terminology issues in user access to Web-based medical information. *AMIA'98.* American Medical Informatics Association.

McLeod. S. D. (1998) The quality of medical information on the Internet: a new public health concern. *Archives of Ophthalmology, 116*, 1663.

Pew Internet & American Life. 2002. Vital decisions: How Internet users decide what information to trust when they or their loved ones are sick. *Pew Internet & American Life Project Report 2002*.

Pew Internet & American Life. 2003. Internet health resources: Health searches and e-mail have become more commonplace, but there is room for improvement in searches and overall Internet access. *Pew Internet & American Life Project Report 2003.*

Reih, S. Y. (2002). Judgment of information quality and cognitive authority in the web. *Journal of the American Society for Information Science and Technology, 53*(2), 145-161.

Saracevic, T. (1975). Relevance: a review of and a framework for thinking on the notion of information science. *Journal of the American Society for Information Science, 26*, (pp. 321 - 343), 1975.

Spink, A., Yang, Y., Jansen, J., Nykanen, P., Lorence, D. P., Ozmutlu, S., & Ozmutlu, H. C. (2004) A study of medical and health queries to web search engines. *Health Information and Libraries Journal, 21*, 44-51.

Tufts. (2005). *Tufts nutrition navigator* [Web site]. Retrieved January 2005, from http://navigator.tufts.edu/comm

Zeman, F. J. (1983). *Clinical nutrition and dietetics*. New York: Macmillan.