

Nutritional Information on the Web: An Analysis of Information Sought and Information Provided

Susan G. Doran and Caroline M. Eastman
Department of Computer Science and Engineering
University of South Carolina
Columbia, SC 29208
803-252-3765, 803-777-8103
sue_doran@yahoo.com, eastman@cse.sc.edu

Bernard J. Jansen
School of Information Sciences and Technology
The Pennsylvania State University
University Park, PA, 16801
814-865-6459
jjansen@acm.org

ABSTRACT

Whether users searching for nutrition related information on the Web 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.

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 Excite and Google, two 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. For the general search engines, matches were retrieved for about 92% of the queries. Most of the matches in both cases were judged relevant.

INTRODUCTION

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 address related work, the methodology of our study and the results. This is followed by a discussion and the implications for user information seeking. We end with directions for future research.

RELATED WORK

A sampling of pertinent statements from recent related work follows. User queries were analyzed in a study due to the suggestion that better design of systems will come from studying user behavior and user perception of IR [Moukdad & Large, 2002]. New tools for searching 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]. In 2002, research by Cothey concluded with a plea for greater understanding of Web information searching so that information could be effectively provided. Doran [2002] provides a more extensive literature survey.

RESEARCH METHODOLOGY

This study utilized seventy nutrition queries submitted to 13 nutrition sites. Excite, and Google. 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.

Research Question

Our research objective was to determine to what extent nutritional sites are providing the information that "general" web users are seeking.

Query Selection

The 70 queries used (see Table 1) are from an Excite transaction log containing actual user requests. 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.

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, 2002]. Tufts uses a twenty five-point scale to rate the sites: twenty points for content and five for usability. A stratified sample was chosen with some sites having high ratings and others low ratings.

Study Design

Each query was copied and pasted into the search area of each nutrition site and search engine. Three users reviewed the resulting matches. A match was relevant if all users agreed that it answered the query. The three users were in agreement 100% of the time. Overall results are in Table 2.

RESULTS

Discussion

Detailed analysis of query performance by sites is given by 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. The two search engines retrieved a higher number of topical matches; Google retrieved matches to 67 queries, and Excite retrieved matches to 62.

Table 1: Top 70 Nutritional Queries

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 cal's 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
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	'physiological 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 content	12

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 by 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

Table 2: Results for All Web Sites

Table 2: Results for All Web Sites					
Website/Search Engine Name, Address	Tufts Rating (Out of 25)	Total # Queries Retrieving Matching Results	Percent Matches	Total # Relevant Matches	Percent Relevant Matches
Google, http://www.google.com	N/A	67	95.7%	67	100%
Excite, http://www.excite.com	N/A	62	88.6%	62	100%
Food and Drug Administration, http://www.fda.gov/oc/food.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%
Atkins Nutritionals, http://atkinsnutritionals.com/index/	12	25	35.7%	25	100%
Health World Online, http://www.healthworld.net	12	25	35.7%	25	100%
Food, http://lancasteronline.com/food	22	24	34.3%	24	100%
Consumer Information Center, http://www.consumerinformation.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.thedietchick.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	92.21%	64.5	100%
Standard Deviation: Search Engines Only	N/A	3.54	5.0%	3.54	0%
Average: Nutrition Sites Only	18.6	25.84	35.5%	24.9	90.9%
Standard Deviation: Nutrition Sites Only	4.4	11.7	16.7%	35.5	48%

the American Cancer Society retrieved only 38 matches, of which 33 were relevant. 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. Of note 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.

CONCLUSIONS

The average percentage of queries answered by the nutrition sites and search engines was 35.5% and 92.1%, respectively. If the Excite queries compare to those typically made to these sites, then improvements to the IR systems and content of the nutritional sites are needed. Based on the high percentage of query matches supplied by Excite and Google, users have a better chance of locating relevant information using general search engines rather than using niche sites. Current nutrition information seekers apparently have a choice between small, controlled sites without enough information and large sites with uncontrolled information.

Future work will include further analysis of the existing data, such as the relationship of site focus, if not nutrition in general, to the results, the relationship of the Tufts rating to the retrieval results, and possible 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.

REFERENCES

- Aguillo, I. (2000). A new generation of tools for search, recovery and quality evaluation of World Wide Web medical resources. *Online Information Review*, volume 24 (2). Retrieved February 15, 2002, from Emerald database <http://pippo.emeraldinsight.com>.
- Cothey, V. (2002) A longitudinal study of World Wide Web users' information- searching behavior. *Journal of the American Society for Information Science and Technology*, 53 (2). Retrieved February 15, 2002. from Science Citation Index Expanded Database <http://www.Webofscience.com>.
- Doran, S. G. (2002). Nutrition information retrieval by web users. MS thesis. Department of Computer Science and Engineering. University of South Carolina. Columbia, South Carolina
- 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.
- Moukdad, H., & Large, A. (2001). Users' perceptions of the web as revealed by transaction log analysis. *Online Information Review*, 25 (6). Retrieved February 15, 2002. from Emerald database <http://pippo.emeraldinsight.com>.
- 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)321—343, 1975.
- Tufts. (2002). *Tufts nutrition navigator* [Web site]. Retrieved April 2002. from the World Wide Web:<http://navigator.tufts.edu/com>.
- Zeman, F. J. (1983). *Clinical Nutrition and Dietetics*. New York: Macmillan.