An Analysis of Cooking Queries: Implications for Supporting Leisure Cooking

Sally Jo Cunningham
Computer Science Department
University of Waikato
Hamilton, New Zealand
sallyjo@cs.waikato.ac.nz

David Bainbridge
Computer Science Department
University of Waikato
Hamilton, New Zealand
davidb@cs.waikato.ac.nz

Abstract

Cooking is a common and an information-intensive activity. We analyze a set of 678 cooking-related queries to identify the attributes that cooks provide in their queries to the Google Answers™ 'ask an expert' online reference system. The results suggest directions to take in developing an effective organization and improved functionality for a cooking-focused digital library.

Keywords: cooking, information needs, query analysis

Introduction

Cooking is a common, complex, information-intensive activity. The traditional information resource is the physical cookbook (delightfully referred to by Teng et al, 2011 as "the tombs of printed recipes … found in almost every kitchen"). Recipe and other cooking-related documents are widely available on the Web, and the support of cooking through improved access to cooking information is an active and growing area of research.

However, at present significant gaps exist in our understanding of information seeking behavior in regards to cooking: What are leisure (that is, non-professional) cooks looking for when they consult cookbooks, websites, cooking videos, etc.? How do they describe their information needs? What needs do they find difficult to fulfill? Answers to these questions may point to enhancements to current digital cooking resources—additional access points, functionality, and document and information types. This paper adds to our understanding of cooking information behavior through an analysis of a significantly sized (678) set of cooking questions posed to an online community reference system (Google Answers).

This paper is organized as follows: we first explore related work into cooking information needs and how they can be supported (RELATED WORK); we next describe the source, characteristics, and analysis method for authentic cooking information requests examined in this present paper (DATA GATHERING AND ANALYSIS); we explore the broad categories of information that the posters wanted (CHARACTERIZING THE DESIRED RESPONSE); we closely examine how posters describe one of the most commonly requested categories, the recipe, and suggest possibilities for enhancing recipe searching / browsing (DESCRIBING RECIPES); and we conclude by considering how a full-fledged cooking digital library could move beyond the bare-boned functionality of a recipe collection (SUPPORTING COOKING IN A DIGITAL LIBRARY).

Related Work

The work related to this study falls into three categories: ethnographic investigations of cooking information needs; examination of recipes and their interpretation; and examining usage of software that supports cooking.

Cunningham, S. J., & Bainbridge, D. (2013). An analysis of cooking queries: implications for supporting leisure cooking. *iConference 2013 Proceedings* (pp. 112-123). doi:10.9776/13160

Ethnographic Studies of Cooks and Cooking

A rich literature exists describing cooking related information behavior of serious but non-professional cooks (Hartel, 2006; Hartel, 2010a; Hartel, 2010b). These studies were conducted as small scale (20 to 31 participants) interviews and observations of self-identified gourmet / hobbyist cooks, with a focus on uncovering the time-based behaviors and processes in which cooking information behavior is embedded (Hartel, 2006; Hartel, 2010b), the management of personal cooking information collections (Hartel, 2010a), and the resources hobbyist and gourmet cooks consult (Hartel, 2006; Knopp, 2011). An earlier examination specifically of cooking information needs of hobbyist cooks (Knopp, 2011) relied on 24 descriptions of 'cooking episodes' elicited from study participants. In that study, the primary information requests were for recipes, explanations of cooking techniques, and where to obtain hard-to-locate ingredients. Through this previous research, a rich picture has been developed of the sources consulted by 'serious leisure' cooks (Stebbins 2007), the ways that they consume this gathered information, the points in their cooking-related activities at which they seek further information, and how they build and maintain their personal information collections.

Less is known about information behavior and needs of novice cooks or to support pedestrian, 'everyday' home cooking. Van de Wiel et al (2010) explore the attitudes to cooking, cooking habits, and cooking / eating contexts of young male professionals with low levels of cooking skills, through contextual interviews of 11 participants. Key findings motivated feature development to the Cook*King* system, including supporting social and competitive aspects of dish preparation and adding mobile functionality to aid in ingredient shopping. A diary study of five young adults with limited cooking experience (Palay & Newman, 2009) elicited similar results, and also identified the need for assistance in achieving greater variety in meals.

These studies adopt intensive qualitative methodologies (diary study, interviews, observations), and so necessarily are based on limited numbers of participants. This present paper examines a significantly larger data set (678 cooking related queries posted to the Google Answers forum)—but far less background and contextual information is available for these queries than with the earlier studies. For example, the Google Answers queries differ from the information needs as described in (Knopp, 2011) in several significant ways: the postings are not necessarily tied to a specific cooking event (that is, a query could be part of an ongoing information gathering exercise and was not necessarily tied to creating a particular dish or meal), and the Google Answers posters themselves had not been able to satisfy their information need (whereas the participants in (Knopp, 2011) were asked to describe information needs for a successful cooking event, and so presumably their questions had been answered).

These earlier studies targeted participants with a common cooking skill level (eg, gourmet hobbyists, young men with limited cooking experience). By contrast, the queries from this present study span the spectrum from confident gourmet hobbyists (eg, "I'd like to make high-quality sushi at home", to the cautious experimenters (eg, "How do you make Hamburger Helper-Cheeseburger Macaroni from scratch?"), to utter novices:

> I'm a 19 year old guy living in Singapore with nothing to do for 2 weeks. I am thinking of baking/cooking for fun! What should i know/have before i start on my adventure? What are the tools involved? How do I bake a cookie?

Examining Recipes and Their Interpretation

Another approach to supporting information behaviour in cooking begins with a close examination of information presentation structures in recipes, and how people understand (or misunderstand) them.

Hamada et al (2000) and Shirai & Ookawa (2006) analyse collections of Japanese recipes to develop models of the steps involved in cooking, with the goal of automated processing existing recipes to produce multimedia / animated 'walkthroughs' of how to cook a given recipe. Unacceptably high rates of errors in automated extraction of cooking actions arise from variation in vocabulary across recipes—highlighting the difficulties people face when learning to cook. One aspect of this vocabulary problem is examined by Yamakata et al (2009): how people naturally refer to ingredients in intermediate stages of preparation (eg, "minced onion"). Here a surprising variation in vocabulary of the participants was identified, compounding the variation in recipe vocabulary noted by the earlier research. Though this was a small-scale study (involving 20 participants), its results are promising in suggesting query processing techniques to match a searcher's native vocabulary to existing recipes, over all steps in a recipe.

Van Pinxterin et al (2011) investigate how people perceive similarity in recipes, through a card sorting study (14 participants, sorting 66 recipes). A model was induced of the most significant characteristics contributing to perceived similarity, and this model was incorporated into a cooking support system that recommended 'healthy' dishes similar to earlier recipe selections of the participants. The authors note that their recipe model was 'quite different from other approaches in the literature'.

These studies attempt to leverage off the large number of existing, digitized recipes, to identify techniques to make these recipes more useful, usable, or accessible (particularly to novice cooks). Further research on human perceptions of recipe structure and vocabulary is needed to more effectively support recipe searching and the searcher's understanding of recipe search results; matching between the 'pain points' identified through qualitative explorations of information needs may be useful in guiding this research. However, this exclusive focus on recipes misses out on the prospect of supporting behaviors associated with other significant cooking related documents used by cooking hobbyists (eg, culinary magazines, cooking videos, menus).

Learning from Deployed Cooking Information Systems

A third approach to investigating cooking information behaviour is to examine how existing cooking document collections are used, for example through usage log analysis. Despite the long-standing and widespread availability of large, digital recipe collections, no studies have been conducted to date of how people search and browse these collections—a serious gap (and surprising) gap in the research literature.

However, evaluations of novel features implemented in prototype cooking information systems provide insight into promising directions for further investigation. Users of the Cook*King* system (van de Wiel et al, 2010) found their individual histories of favorite recipes to be useful in meal planning and developing shopping lists. A small-scale case study indicates that an individual's recipe browsing and cooking history can also serve as the basis for a model of their preferences, to inform personalized recipe recommendations (Ueda et al, 2011).

Evidence of other users' preferences, opinions, and recommendations can be useful when searching and browsing for recipes. A six month study of the experimental Kalas social navigation system indicates that users both liked these "social trails" through a recipe collection, and also modified their searching/browsing through Kalas to take advantage of them (Svensson et al, 2005).

Data Gathering and Analysis

The source of the natural language cooking-related queries analyzed in this paper is Google Answers' 'ask an expert' service (*http://answers.google.com*). In this service past queries and their answers are publicly available, arranged into categories (with the person who posts the query selecting the appropriate category for it). This is an attractive source of 'real' questions, and allows us to study active information needs rather than post-hoc recollections (as in Knopp, 2011).

The questions posted to Google Answers are in natural language, describing the information need rather than being restricted to a few user-selected query terms (as is the case in most digital libraries and search engines). For example, an analysis of approximately 15 million MSN search queries found that 90% of queries include four terms or fewer, and 99.9% of queries include twelve terms or fewer (Bendersky and Croft, 2009). In contrast, the Google Answers queries average 71 words. The Google Answers user places a monetary value on the question (the bounty offered was limited to be between US $2–200) and pays the expert on receipt of a satisfactory answer; these are thus significant rather than casual questions. The complete set of the Food and Cooking category to Google Answers™ 'ask an expert' system (*http://answers.google.com*) was downloaded: 678 queries posted to the site from 10 April 2002 to 23 November 2006 (when the service was discontinued). The questions posted to Google Answers are in natural language, describing the information need rather than being restricted to a few user-selected query terms. Queries averaged 71 words.

The complete set of the Food and Cooking category to Google Answers was downloaded; the set totaled 678 queries posted to the site in the period 10 April 2002 to 23 November 2006. The value of postings ranged over the full limit the site allowed with a mean price of US $11.62 (standard deviation of 18.3), and mode and median value being the same, at US $5. Of the 678 questions posted, 49% of them

were successfully answered, meaning the originator of the question accepted the answer, and paid the fee.

Figures 1, 2, and 3 provide a more detailed overview of the pattern to the posed questions and answers, and are useful in determining clusters and extremities in the data.

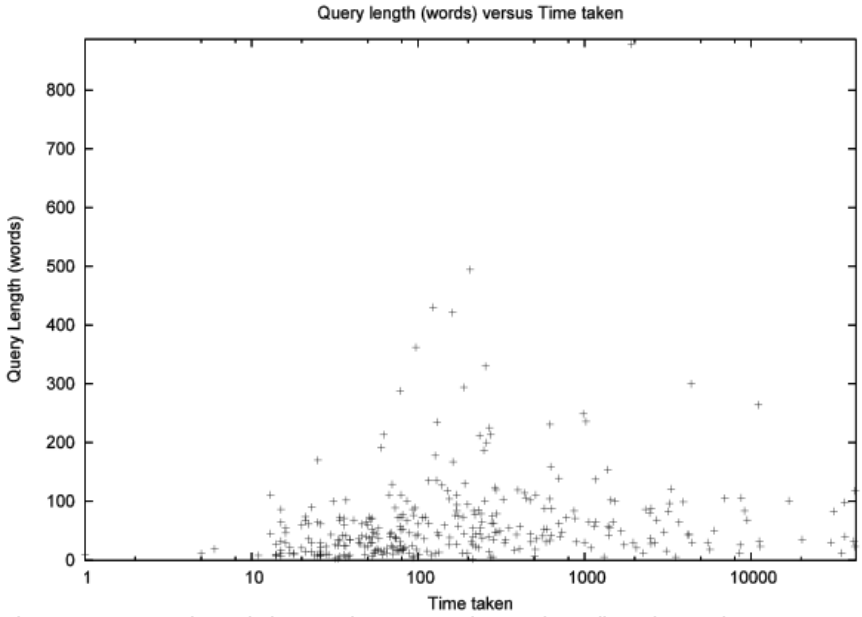

Figure 1. Query length in words versus time taken (in minutes) to answer the query.

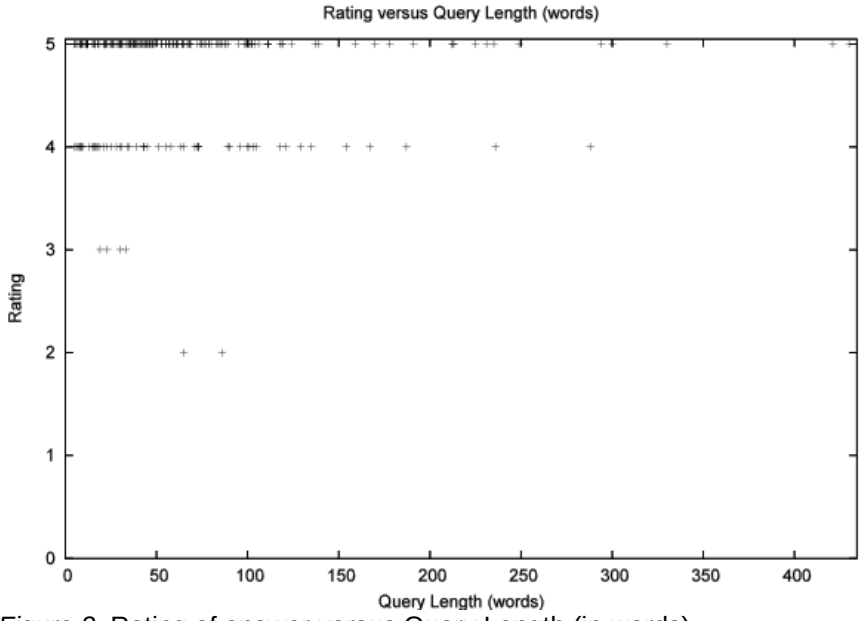

Figure 2. Rating of answer versus Query Length (in words).

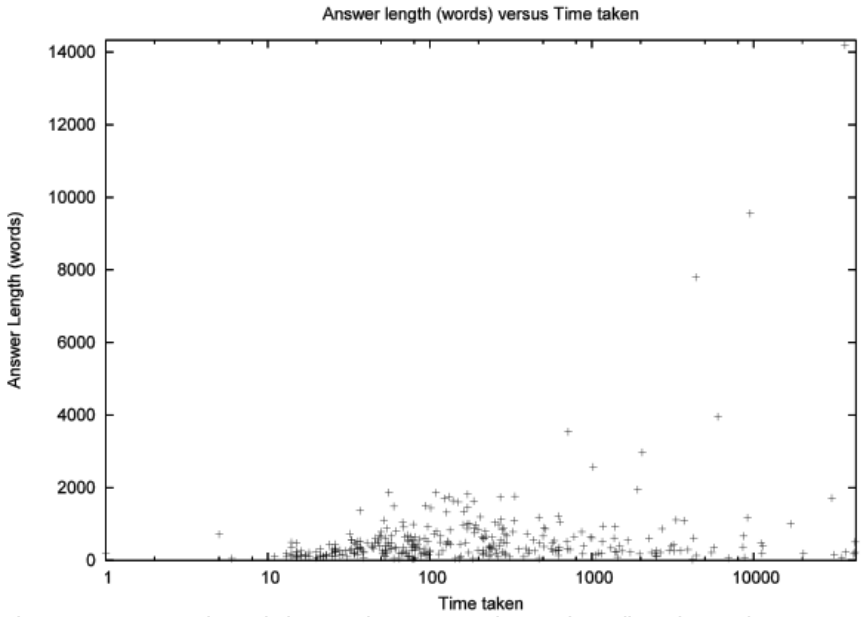

Figure 3. Answer length in words versus Time taken (in minutes) to answer the query.

In Figure 1 we get a sense of how long a user needs to wait to get an answer. Almost half (49%) are answered within the first two hours, rising to 63% answered within the first 4 hours, and 85% within 24 hours. Through this 2-4 hour band we see higher activity in the y-axis (length of query), showing that taking the time to provide a posting with some detail pays dividends in terms of getting a timely answer. The quickest response time was one minute for a question asking about the shelf life of eggs in a refrigerator.

Google Answers posters are given the opportunity to rate the quality of the answers provided. As Figure 2 shows, most users are highly satisfied with the answers given, rating the answer either with the highest or second highest rating. Some lower ratings were assigned to answers where the questions asked themselves were rather brief (under 100 words) compared to others.

Turning our attention to the answers given, Figure 3 visualizes the average number of words in the response to a query against the speed of response. The norm for answers is under 2000 words, with a rising tail indicating that longer answers (not surprisingly) tend to take more time to compile. The longest query was over 14,000 words long. This was a response to a question asking for a list of the top 1000 recipe books based on sales. From this phase of the analysis, 93 of the 678 queries were discarded as being off-topic to this present study of hobbyist cooks: for example, queries about cutlery, china, and glassware ("Where can I can definitely buy this whiskey / whisky glass? http://tinyurl.com/7kmna", queries about commercial equipment ("I am looking for information about how commercial soft serve ice cream machines function, in technical terms."), or advice on starting a food-related business ("Mum wants to sell a little of her home-made cakes, to local shops, in Australia?").

A grounded theory approach (Glaser & Strauss, 1967) was taken to analyze the remaining 586 queries. The queries were coded to characterize the desired response to their query (eg, a recipe, an explanation of a cooking technique). Recipe-oriented queries were further coded to characterize the details that users provided when describing the desired recipe (existing recipe ontologies (Dale, 1989) and (Kimura et al, 2008) were consulted to suggest labels for emergent categories). The coding scheme is described in the RESULTS section.

Characterizing the Desired Response

As summarized in Table 1, the Google Answers posters were seeking a broad range of responses:

Table 1
Breakdown of Desired Response Types

Category	Count	%
Product Information	213	36.41%
Recipes	119	20.34%
Hygiene, Health, Food Storage & Preparation Advice	53	9.06%
Food Trivia and Market Research	42	7.18%
Definitions	39	6.67%
Food History and Background	37	6.32%
Resource	34	5.81%
Troubleshooting	27	4.62%
Menu and Food Combination Suggestions	23	3.93%
Cooking Techniques	21	3.59%
Miscellaneous	20	3.42%
Measures, Conversions, and Substitutions	15	2.56%
Contacts for Experts	5	0.85%

Note: Does not sum to 100%; some queries include more than one desired response category.

Product information. Postings request information about a food item or physical equipment involved in cooking.

Food items ran the gamut from the 'gourmet' ("'Les Langues Dorees' Demel Vienne chocolates") to the humble ("frozen limeade concentrate"). These edible items might be ready-to eat ("Nutria jerky"), an ingredient the poster anticipates to use in his/her own cooking ("sausage skins for home-made sausages"), or a sauce or other accompaniment to a dish ("Looking for how to get a salad dressing named chateau classique. Maybe distributed by Sexton Foods, Inc.").

With food items, posters wished to purchase products they were finding difficult to source—for example, a treat from one's country of origin ("Where can I buy LUCKY CHARMS breakfast cereal in the UK?), a specialist version of a common product ("Looking for fresh farm lamb and other quality meats such as farm pork, farm free range chickens with no growth hormones and antibiotics"), or a commercial product for which no brand name was known ("How can I obtain Powdered alcohol in the united states?"). Less frequently (four queries), if the item itself could not be located then posters requested a substitute ("I had a liquor in Curacao named Ponche Kuba…If it isn't available what other liquor can be substituted for it instead?").

The majority of equipment related posts were seeking to purchase the item. The types of equipment desired included replacement parts, specialist food service and prep gadgets, ovens, and food storage devices. In contrast with the food item requests, fewer of the equipment requests included a brand name of the desired item—the poster instead provided a more-or-less elaborate description of the desired item. The descriptions ranged from the straightforward request for a generic instance of an item ("Where can i buy a JELLY MOLD in toronto?") to detailed requirements:

> I am looking for a recommendation on a blender to bring to work to mix up shakes. Being that this is a typical office environment I need something that won't disturb my coworkers.
> Requirements:
> 1) Super Quiet
> 2) It is for making shakes with ice and water. (needs to be able to crush ice)
> 3) Would be nice if it were the one piece types that are easier to clean. (not entirely one piece, just the top glass portion)
> 4) Looks are unimportant
> 5) Order online for less than $100
> Clarification: I'm looking for a standard blender, not a hand-help model.

Of the equipment queries not requesting purchase information, three queries related to safety or hygiene issues with cooking gear (eg, “is it wrong to run microwave ovens with nothing in them and if yes, why”), and three requested plans or advice in constructing a specialist cooking device (a brick oven, an elaborate wooden outdoor pig roaster, and an apartment-based sake brewing environment).

Recipes. Recipes are requested for complete dishes (eg, “green chilli mashed potatoes”), a sauce or other accompaniment to a meal (eg, “Japanese curry sauce”), a food item that can be used as an ingredient in other recipes (eg, “Mozzarella cheese”), and for common sub-parts of more complex recipes (“A good old-fashioned pie pastry recipe required, savoury/meat etc, not sweet!!”). Recipe queries can further be loosely divided into requests for ‘a’ recipe (that is, any recipe meeting the conditions of the query; “a recipe for Halva”), and requests for ‘the standard’ or ‘the best’ recipe for a specific dish. How posters described recipe requests is examined in the next section, DESCRIBING RECIPES.

Hygiene, Health, and Food Storage & Preparation Advice. These queries include requests for good practice for food handling (“I am a cheapskate, with a lot of uncooked meat which is probabaly [sic] not "bad" yet. But I am not certain. If I freeze this meat, will it kill any "bugs" or bacteria already in the meat? Or, am I simply delaying my trip to the Tomaine Temple?”) and background on potential health effects of food preparation techniques (“Does microwaving food in plastic containers or plastic cling wrap release harmful chemicals into the food?”).

Food Trivia and Market Research. These queries were directed most frequently at ‘top ten’ type lists (“I am looking for the top ten countries in the world in terms of the consumption of clams”) and statistics on purchase and dining trends (“where can i find statistics on whether people leave or drink the milk after finishing all the cereal in the bowl?”; “there are many different kinds of miso in the stores. How do I know which is most like the miso commonly served in Japanese restaurants in North America?”).

Definitions. Posters request: an ‘official’ name for a specific food item or technique (“A raw egg swallowed in Sherry and Worcester Sauce: how is this hangover cure better known?”); a description or definition of a food item (“I was in the indian grocer this afternoon, and a bunch of people were lined up waiting for "tawa". What is "tawa"?”), and an explanation of the difference between two related food items (“What are the differences between Caramel and Butterscotch?”).

Food History and Background. These posts include general requests for ‘history’ of a food item or technique (“What is the history of teppanyaki cooking?”), as well as more specific inquiries about origins and practices (“Afghan Biscuits are a popular biscuit/cookie in New Zealand. How did they get their name?”).

Resources. Posters requested print or digital resources that they could use to satisfy further personal information needs: for example, cookbooks, recipe websites, discussion forums, and more specialist cooking and food related websites (eg, “Strawberry themed cake images”; “a webpage that lists all produce that supermarkets sell, complete with pictures”; “a website that will design a menu based on a few budget friendly ingredients so that you get the most bang for your buck.”).

Troubleshooting. Posters requested information on: how to recover from a cooking disaster (“What takes the burnt flavor from rice when you burn it?”); the underlying causes for a cooking problem experienced in the past and how to avoid them in future (“What mistake did I make when I baked challah bread that was too dense? I followed the recipe perfectly and removed the bread a few minutes before the allotted time.”); and tips for achieving more professional cooking results at home (“How do you cook green vegetables and have them stay green? My green beans always get pale, yet I have had vivid green beans in restaurants.”).

Menu and Food Combination Suggestions. These posts request suggestions on appropriate items to create or complete a specific menu (“I want to make a list of standard spice combinations, correlated with the types of dishes the spice combination works with”; “What foods go with Chianti Wine”), or request more general advice on food and ingredient combinations. (“Do things such as honey, coconut

and the like go well with chicken? What other sweet things can be used to compliment a good piece of chicken?").

Cooking Techniques. These posts request an explanation or description of a cooking technique or food preparation, rather than a recipe per se ("I'm trying to discover the secret to getting steaks to come out the way they do in restaurants if you order them medium rare, with a smooth dark brown crust on the outside but medium rare (light pink throughout the inside.)"). Most technique references are to everyday home cooking situations ("What is the best way to boil a perfect hard-boiled egg?"; "I would like suggestions on how to peel a butternut squash without becoming an amputee."), but some are more specialized and idiosyncratic ("In an Episode of Father Ted Mrs Doyle gives a cake to Eoin MacLove with a sweatshirt inside it. How do I make a similar cake with a sweatshirt baked in the middle of it?").

Miscellaneous. Queries that do not fit into any other category ("I am thinking of baking/cooking for fun! What should I know/have before I start on my adventure?").

Measures, Conversions, and Substitutions. Conversion queries include present day measures ("What's is the approximate volume of 250 g wheat (or rye) flour?") as well as antiquated measures ("I have a recipe for Devil's Food Cake from my great grandmother that calls for '1/3 cake chocolate', and I want to know how much powdered cocoa or unsweetened baker's squares to substitute."). This category also includes queries relating to the calories or dietary units in food items and substitution of food items (eg, Weight Watchers diet 'points') and substitution of food items ("Can red and yellow saffron be used interchangeably?").

Contacts for Experts. These queries are satisfied by names and / or contact details for human experts in various aspects of cooking: food critics, chefs, and wine experts.

Describing Recipes

In its minimal form, a recipe consists of a title, a list of ingredients, the amount of each ingredient, and a set of instructions for processing the ingredients to prepare the named dish. Here, we analyze the descriptive factors posters use to characterize recipe requests (Table 2).

In many print cookbooks and digital recipe repositories, the most common access points (searching and browsing) to the collection are the *name of the dish*, the main *ingredient*, and the *course or meal* associated with the dish. These three are the most frequently occurring descriptor types in the Google Answers queries. However, specifying these might not be straightforward.

There might be no 'official' name for a dish, or no single standard English spelling for a dish originating in another culture. A dish name may reflect its appearance (eg, Toad in the Hole, an English standard involving sausages in batter) or folktale / cultural associations (The Imam Fainted, a Turkish eggplant dish) rather than its ingredients. These difficulties are common enough to suggest that query interfaces should not rely overly heavily on recipe name searching. While prototype systems exist that try to work around difficulties in named dish searching by, for example, matching to query descriptions of a recipe's process structure (Wang et al, 2008), these systems are far from deployment.

On the other hand, abandoning recipe title in favor of simple keyword search over the entire recipe text (as is common in current online recipe collections) can be frustrating as well, when the user wants a recipe for a food item that is commonly an ingredient in other dishes (eg, mozzarella cheese, sausage). In these cases, the option of searching by name / title can be invaluable. The usual difficulties with 'bibliographic' query pertain as well—users' spelling may include errors, and they may misremember the name or other significant metadata for a recipe (eg, "Spelled to sound like "Padiusca"" for *puttanesca*). Spell-checking and browse-able term lists would be useful in addressing these problems.

Table 2
Characteristics Used to Describe Recipes in Queries

Category	Example	Count(%)	Category	Example	Count(%)
Name of dish/item	Raisin tea cake	67 (56%)	Dietary restrictions	Vegetarian, gluten free, no dairy	15 (13%)
Ingredient	Bean sprouts, meat, roasted chicken	64 (54%)	Cooking technique	Dry curing, barbecue	15 (13%)
Type of dish, course, meal	Pasta, appetizer, breakfast	34 (29%)	Similarity to an example	Similar to a Mounds bar	15 (13%)
Ethnic/National/Region	Japanese, UK, Montreal	29 (24%)	Difficulty level	easy	12 (10%)
Commercial, restaurant, chef	Coco Ichiban Curry House, George Shearing	29 (24%)	Print/Web resource where 1st located	Betty Crocker Creative Recipes: 1995-1997	10 (8%)
Time period or date	1950s, a year ago, ancient	22 (18%)	Holiday, event, special occasion	Valentine's Day, tailgate party	6 (5%)
Sensorial aspects	Crispy, buttery, flakey	21 (18%)	Procedure	Cut it up and boil it	5 (4%)
Appearance	Yellow, round, tapered	16 (13%)	Cost	Cheap, cost-effective	3 (3%)
Locally available ingredients	Available in UK	16 (13%)			

Note: Does not sum to 100%; most queries include more than one information need descriptor.

Ingredients can be tricky to specify: can I say 'meat' or must I enumerate all locally available possibilities ('beef or chicken or lamb or pork')? Does the recipe use American terminology or English (eg, 'bell pepper' vs. 'capsicum')? Prototype systems presented in (Kimura et al, 2008) and (Liu et al, 2005) demonstrate the feasibility of ontology support for queries including ingredient specification.

Queries seeking recommendations for suitable dishes to serve frequently specify a *course or meal*, a *holiday/event/special occasion*, *ingredients* that the cook wishes to use (or use up), or degree of *similarity* to another dish that the person has tried. Recipe recommender systems are an active area of research (eg, Teng et al, 2011; Ueda et al, 2011; van Pinxteren et al, 2011) and can incorporate these facets of a need description.

A quarter of queries are for a specific version of a dish: the version created or popularized by a *commercial organization, restaurant, or chef*. These queries typically state that the user will not be satisfied with any other versions of the dish—suggesting that the 'author' or originating organization be included as metadata to recipes when this information is available.

Nearly a quarter of queries specify an *ethnic, national, or regional* origin for the desired recipe. Of course, all dishes come from some cultural context, so the remaining three-quarters of recipes are assumed to pertain to the dominant nationality of the Google Answers users (USA). The cultural assumptions for the target audience of a recipe (as distinct from the origin of a recipe—eg, a Korean recipe written for an American audience) should be made clear, as the measurements, names for ingredients, and common cooking techniques can vary across cultures (Kimura et al, 2008).

Similarly, posters requesting a recipe whose origins did not match the poster's current physical location recognized that they might run into trouble sourcing ingredients, and so specified that they wanted recipes only with *locally available ingredients*. It is not too far-fetched to posit a system that could match a user's current location to the inventory list of local online grocery stores, and use that information to remove or down-rank candidate recipes whose ingredients would be difficult to source.

Time period appears most useful when the poster is seeking historic recipes (eg, "ancient and traditional methods for curing meat") or when tied to an iconic menu of a well-known *restaurant or chef*. Precise dates would be overkill—these queries are better served by broad categories such as 'Middle Ages' or '1950s'.

Sensorial aspects of desired dish (including texture, 'mouth feel', and intensity of flavor) address the impact of a dish on the diner—and so may be particularly useful in selecting between alternative recipes for a given dish. The system described in (Liu et al, 2011) supports search refinement by sensorial aspects by reasoning from a large-scale database of sensorial facts about ingredients.

The *appearance* of the finished dish is most frequently (and most elaborately) described when the poster hopes that something distinctive in the desired dish's form will assist a respondent in identifying it, or allow the respondent to distinguish between similar dishes. Providing accompanying photos of finished dishes to recipes could be useful for these posters, as an aid in browsing potentially relevant recipes.

Current recipe systems put the onus on the creator or recorder of the recipe to indicate whether a specific recipe meets special *dietary restrictions*. Given that many such restrictions are defined by the elimination of a category or type of ingredient from recipes (eg, vegetarian, no dairy), it is not far-fetched to imagine a system that can automatically infer whether a recipe meets the standards for common diet regimes—and indeed, the system described in (Liu et al, 2005) prototypes this facility. Less common restrictions may be supportable by negations in queries (eg, one poster's son found cream cheese to be indigestible—'not cream cheese').

Queries may include the *cooking techniques* and *procedures* for a dish when the poster does not know the dish's name. Search over these facets can be supported by 'mining' the instructions for a recipe to create a representation of its workflow, and then comparing that to the query workflow (Wang et al, 2008).

When *difficulty level* is specified, posters ask for a simple, easy to prepare recipe that requires no specialist equipment—suggesting a binary rating of difficulty rather than more elaborate schemes. Similarly, the few queries specifying *costs* uniformly ask for cheapness, rather than finer-grained cost ratings.

Posters include surprisingly precise details of the *print or Web resource* in which a 'lost' recipe was originally located. In the case of print resources, the difficulty is generally that the document has gone out of print—a problem that likely will be eased as more documents are digitized (possibly as individuals violate copyright to post favorite recipes). Links to Web recipes may break as the websites disappear or the site re-organizes, pointing to a need for personal digital collection support (Hartel, 2010a).

Supporting Cooking with a Digital Library

The above analysis clearly shows the needs of the Google Answers posters are wide-ranging. In terms of how a digital library infrastructure could help support their cooking information needs, our analysis suggests that the identified activities map well to a managed set of digital library collections, where cross collection searching and browsing the "super-structure" that matches the categories developed through ground truth analysis is supported. In addition the digital library environment should be augmented with embedded analytical tools, and support social structures such as forums and blogging.

For instance, the *Food History and Background,* and *Recipes* collections could be populated by the aggregation of existing on-line resources dedicated to cooking. If a protocol such as OAI-PM (Buchanan et al, 2005) were to be used, then these distributed resourced could be regularly scanned for new and corrected content, and updated accordingly. Such co-operation of the primary source providers seems likely if—having located an item through the aggregated central collection—the user is taken to the original site so due recognition (and any relevant revenue stream) is given. If such an initiative were supported by Amazon (with their "look inside" capability) the *Resources* collection could be developed along similar lines, helping bridge the gap between the physical and digital forms hobbyist cooks are known to draw upon. Similarly a *Product Information* collection could gain significant leverage from integration with shopping aggregator sites such as Google Products.

The *Cooking Techniques* collection would work well as a video collection with content sourced from sites such as YouTube or commercial content providers. The *Definitions* collection could be seeded by text mining for facts in the other collections in the digital library. Tools embedded in the digital library environment would be essential, none more so than a measurement conversion capability. Coping with domain specific esoteric or dated forms of measurement that come up from time to time would be an interesting challenge—perhaps another opportunity for text mining to assist, this time by being applied to the *Food History and Background* collection.

Supporting *Menu and Food Combination Suggestions* is perhaps the biggest reach in terms of what is available in off-the-shelf digital library software. However, existing recipe recommender systems (eg, Teng et al, 2011) could be integrated into a recipe digital library.

Forum(s) for discussion would fill out the user experience. Our analysis suggests there is a niche for a specialist *Trouble Shooting* area where people can post and expect a speedy response.

We also noted earlier that hobbyist cooks also maintain a personal collection of resources, and that usage analysis of the CookKing (van de Wiel et al, 2010) and Kalas (Svensson et al, 2005) prototype systems indicate that users value personal histories of favorite recipes. A final "step-up" to what the digital

library provides, then, is to provide registered users with a personal space within the system. Given the approach details, this would be most naturally accomplished as a personal digital library collection. Using the idea of Digital Library Talkback (Bainbridge & Witten, 2011) a two channel can be formed: documents located by a user in the main (publicly available) collections can be transferred to their private collection; and counterpoised to this, content developed within a user's private collection can be exported (published) and included in the publicly available collections.

Conclusion

By analyzing authentic queries to Google Answers, we have elicited a rich framework suggesting the broad range of responses that people want when they search for cooking information, and the types of metadata that they are able to provide when searching specifically for recipes. These posts represent user needs unfiltered by expectations of how a retrieval system expects a query to be presented, the access features that it supports, or the documents it contains. This helps to establish a clearer picture of what users really want and need in a cooking information system.

References

Bainbridge, D and Witten, I.H. 2011. Perambulating libraries: demonstrating how a Victorian idea can help OLPC users share books. In *Proceedings of the 11th Annual International ACM/IEEE Joint Conference on Digital libraries* (pp. 471-472).

Bendersky, M. and Croft, W.B. 2009. Analysis of long queries in a large scale search log. In *Proceedings of the 2009 Workshop on Web Search Click Data* (pp. 8-14).

Buchanan, G., Bainbridge, D., Don, K.J., and Witten. I.H. 2005. A new framework for building digital library collections. In *Proceedings of the 5th ACM/IEEE-CS joint conference on Digital libraries* (pp. 23-31).

Dale, R. 1989. Cooking up referring expressions. In *Proceedings of ACL '89 (Association for Computational Linguistics)* (pp. 68-75).

Glaser, B., and Strauss, A. (1967) *The discovery of grounded theory: Strategies for qualitative research.* Chicago: Aldine de Gruyter.

Hamada, R., Ide, I., Sakai, S., and Tanaka, H. 2000. Structural analysis of cooking preparation steps in Japanese. In *Proceedings of the fifth international workshop on Information retrieval with Asian languages* (IRAL '00) (pp. 157-164).

Hartel, J. 2006. Information activities and resources in an episode of gourmet cooking. *Information Research* 12(1) (October 2006), paper 281.

Hartel, J. 2010. Managing documents at home for serious leisure: A case study of the hobby of gourmet cooking. *Journal of Documentation,* 66*(*6), 847-874.

Hartel, J. 2010. Time as a framework for information science: Insights from the hobby of gourmet cooking. *Information Research,* 15(4) (Dec. 2010).

Kimura, M., Kitamura, Y., Matsuda, M., and Tijerino, Y. 2008. English-Japanese cooking recipe translation system using ontology. *IEICE Technical Report: AI2007-4*, vol 107, 22-82.

Knopp, M. 2011. Information needs, preferences, and behaviors of home cooks. *Library and Information Research,* 35(109), 40-54.

Liu, H., Hackenberry, M., and Selker, T. 2011. Synesthetic recipes: foraging for food with the family, in taste-space. In *ACM SIGGRAPH 2005 Posters*, Article 12.

Palay, J. and Newman, M. 2009. SuChef: an in-kitchen display to assist with "everyday" cooking. In *Proceedings of the 27th International Conference Extended Abstracts on Human factors in Computing Systems* (CHI EA '09) (pp. 3973-3978).

Shirai, K. and Ookawa, H. 2006. Compiling a lexicon of cooking actions for animation generation. In *Proceedings of the COLING/ACL Conference Poster Sessions* (COLING-ACL '06) (pp. 771-778).

Stebbins, R.A. 2009. Leisure and its relationship to library and information science: bridging the gap. *Library Trends,* 57(4), 618 – 631.

Svensson, M., Hook, K., and Coster, R. 2005. Designing and evaluating Kalas: A social navigation system for food recipes. *ACM Transactions on Computer-Human Interaction,* (12, 3) (September 2005), 374-400.

Teng, C-Y, Lin, Y-R, and Adamic, L.A. 2011. Recipe recommendation using ingredient networks. arXiv repository. URL: http://arxiv.org/abs/1111.3919

Ueda, M., Takahata, M., and Nakajima, S. 2011. User's food preference extraction for personalized cooking recipe recommendation. In *Proceedings of the Second Workshop on Semantic Personalized Information Management: Retrieval and Recommendation*. Retrieved from http://iswc2011.semanticweb.org/fileadmin/iswc/papers/workshops/spim/spim2011_paper4.pdf

van de Wiel, E., Meijering, M., and Shahid, S. 2010. Cook*King*: a king of healthy, fun and social cooking. In *Proceedings of the 7th International Conference on Advances in Computer Entertainment Technology* (ACE '10) (pp. 110-111).

van Pinxteren, Y., Geleijnse, G., and Kamsteeg, P. 2011. Deriving a recipe similarity measure for recommending healthful meals. In *Proceedings of the 16th International Conference on Intelligent User Interfaces* (IUI '11) (pp. 105-114).

Wang, L., Li, Q., Li, N., Dong, G., and Yang, Y. 2008. Substructure similarity measurement in Chinese restaurants. In *Proceedings of WWW 2008* (pp. 979-988).

Yamakata, Y., Funatomi, T., Kakusho, K., and Minoh, M. 2009. A study of object naming according to the manufacturing processes in a cooking activity. *Proceedings of the ACM Multimedia 2009 Workshop on Multimedia for Cooking and Eating Activities* (CEA '09) (pp. 1-8).