1-2010

Supporting Multiple Paths to Objects in Information Hierarchies: Faceted Classification, Faceted Search, and Symbolic Links

Saverio Perugini
University of Dayton, sperugini1@udayton.edu

Supporting Multiple Access Paths to Objects in Information Hierarchies: Faceted Classification, Faceted Search, and Symbolic Links

Saverio Perugini

Department of Computer Science
University of Dayton
300 College Park, Dayton, OH 45469–2160, USA

Abstract

We present three fundamental, interrelated approaches to support multiple access paths to each terminal object in information hierarchies: faceted classification, faceted search, and web directories with embedded symbolic links. This survey aims to demonstrate how each approach supports users who seek information from multiple perspectives. We achieve this by exploring each approach, the relationships between these approaches, including trade-offs, and how they can be used in concert, while focusing on a core set of hypermedia elements common to all. This approach provides a foundation from which to study, understand, and synthesize applications which employ these techniques. This survey does not aim to be comprehensive, but rather focuses on thematic issues.

Key words: Faceted browsing and search, Faceted classification, Symbolic links, Web directories, Hypermedia

1. Introduction

Information hierarchies classifying items are prevalent on the web. Hierarchies support both exploratory (e.g., 'What new laptops are available?')

Email addresses: `saverio@udayton.edu` (Saverio Perugini)
URL: `http://academic.udayton.edu/SaverioPerugini` (Saverio Perugini)

Accepted for publication in Information Processing and Management *July 19, 2009*

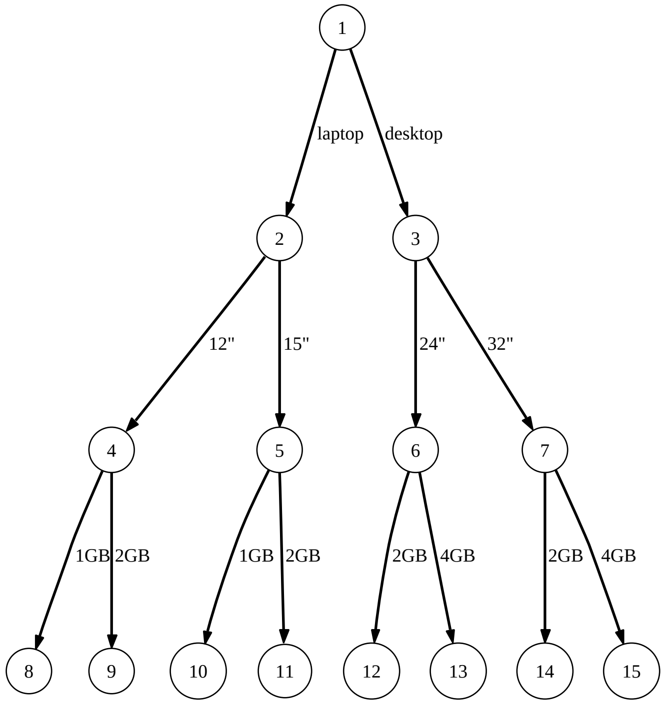

Figure 1: Sample structured web hierarchy simplified for purposes of presentation.

and focused goals (e.g., 'Does a laptop with a 15" display come with 2GB of memory?') through a familiar mode of interaction: browsing. Summarizing and capturing one's information need using keywords as in web search is unnecessary (Baeza-Yates and Ribeiro-Neto, 1999). However, to be effective, the organization of the hierarchy must match or closely match the user's mental model of the task. 'The ability to accommodate different users who may approach the same information from different perspectives is an essential feature for a successful information retrieval' (Glassel, 1998). In this article we progressively study three approaches to support multiple access paths to each terminal object in information hierarchies, and examine their support for accommodating multiple mental models of information seeking or, in other words, different views of the same hierarchy (Botafogo et al., 1992). We start by drawing a distinction between simple hierarchies in their presentation, not implementation.

Fig. 1 illustrates a directed acyclic graph (DAG) model of a web hierarchy with characteristics similar to those one might find in a typical e-commerce site for computer shopping. Vertices represent webpages, directed edges correspond to hyperlinks between pages, and edge labels depict hyperlink labels (i.e., the text between `<a href="">` and `</a>` HTML tags). At the root level, the user is asked to make a choice of *computer type* (laptop or desktop), followed by a selection of *display size*, and, finally, a choice for *amount of memory* (RAM). We call sites like this, where the link labels residing at the same level of depth in the DAG correspond to a *facet* (Wynar, 2000) of information categorization, *faceted*. Facets of a class or subject are 'clearly defined, mutually exclusive, and collectively exhaustive aspects, properties, or characteristics of a class or specific subject. ..., a term introduced into classification theory and given this new meaning by the librarian S.R. Ranganathan and first used in his *Colon Classification* in the early 1930's (Wynar, 2000). We call faceted sites where all leaves (vertices which are the source of no edge) reside at the same level *structured* (Perugini, 2009) because the paths of the site can be modeled by a table in a relational database system (Garcia-Molina et al., 2001). The tree in Fig. 1 models a structured site. A *path* is a sequence of edges from the root of a DAG D to a leaf in D, e.g., $\prec$laptop, 15", 2GB$\succ$ is a path in the DAG in Fig. 1. In such schemes, often an information object is classified in one place and there is a single path to it (i.e., the DAG is a tree). This method of organization closely resembles how objects are organized in the real world (i.e., an object can only be in one place at a time).

Often information hierarchies have dependencies underlying the facet values (also called *labels*). For instance, in Fig. 1 all paths containing an edge labeled '4GB' also have one labeled 'desktop.' On the other hand, none of the paths with an edge labeled '24"' also have one labeled 'laptop.' Each dependency satisfied by a site is captured by a *web functional dependency* (or *web FD*) (Perugini and Ramakrishnan, 2007), a concept related to correlated facets (Ben-Yitzhak et al., 2008). The former dependency is a *positive web FD* (4GB $\rightarrow$ desktop) and the latter is a *negative web FD* (24" $\rightarrow$ $\neg$ laptop) (Perugini and Ramakrishnan, 2007). Any hierarchy with more than one path, by virtue of the mutual exclusive dichotomies implicit in such a hierarchy, must satisfy negative web FD's. However, some hierarchies may not satisfy any positive FD's. It is not possible for any hierarchy to satisfy neither positive or negative FD's. Positive FD's satisfied by any structured site are the analog of database FD's (Garcia-Molina et al., 2001) satisfied by the corresponding relation. On the other hand, the database relation corresponding to a structured hierarchy satisfying no positive web FD's satisfies all possible multivalued dependencies (MVD's) (Garcia-Molina et al., 2001). Informally, each facet is completely independent of all others and, thus, the set of paths in the site correspond to the Cartesian product of all sets of facet values. In other words, each facet value lies along a path with every other value outside of the facet in which it is a member. For instance, in contrast to Fig. 1, a structured site consisting of the $\{\prec$laptop, 12”, 1GB$\succ$, $\prec$laptop, 12”, 2GB$\succ$, $\prec$laptop, 15”, 1GB$\succ$, $\prec$laptop, 15”, 2GB$\succ$, $\prec$desktop, 12”, 1GB$\succ$, $\prec$desktop, 12”, 2GB$\succ$, $\prec$desktop, 15”, 1GB$\succ$, $\prec$desktop, 15”, 2GB$\succ\}$ set of paths satisfies the $\{$computer type $\rightarrow\rightarrow$ display size, computer type $\rightarrow\rightarrow$ amount of memory, display size $\rightarrow\rightarrow$ computer type, display size $\rightarrow\rightarrow$ amount of memory, amount of memory $\rightarrow\rightarrow$ computer type, amount of memory $\rightarrow\rightarrow$ display size$\}$ set of nontrivial MVD's (Garcia-Molina et al., 2001) and satisfies no positive FD's. A structured site where the complete set of terminal objects is described by the Cartesian product of the sets of facets values for each facet, such as that whose sequences are given above, satisfies no positive FD's. In other words, structured sites where any value for any facet can be combined with any value for any other facet satisfy no positive FD's. Moreover, it is impossible for an unstructured, faceted site to satisfy no web FD's because a path which is shorter than another path cannot involve values from any unrepresented facet, and this fact is captured in a negative web FD. The term inference involved in web FD's is related to the more general inference of valid and invalid compound terms in a faceted taxonomy involved in the *Compound Term Composition Algebra (CTCA)* (Tzitzikas et al., 2007).

The number of vertices in a faceted site varies with the top-down ordering of the facets in the hierarchy. For instance, while the hierarchy in Fig. 1 has 15 vertices, a hierarchy with the same terminal objects rendered using a *by display size, computer type, amount of memory* motif contains 17 vertices. Of course, in unstructured, faceted sites, some facet orderings are not permitted. For instance, a facet which is not represented in at least one path cannot reside at the root level, since if it did the terminal object residing at the end of this path would be inaccessible. The number of vertices in a structured hierarchy which is a tree satisfying no positive web FD's is described using the formula $1 + |v_1| + \sum_{i=2}^{|f|} (\prod_{j=1}^{i-1} (|v_j|) \times |v_i|)$, where f is the set of all facets and v_i is the set of facet values for facet f_i. More importantly, there is only one path available to access each terminal object (by the definition of a tree) constituting the interaction limitations of such sites. Therefore,

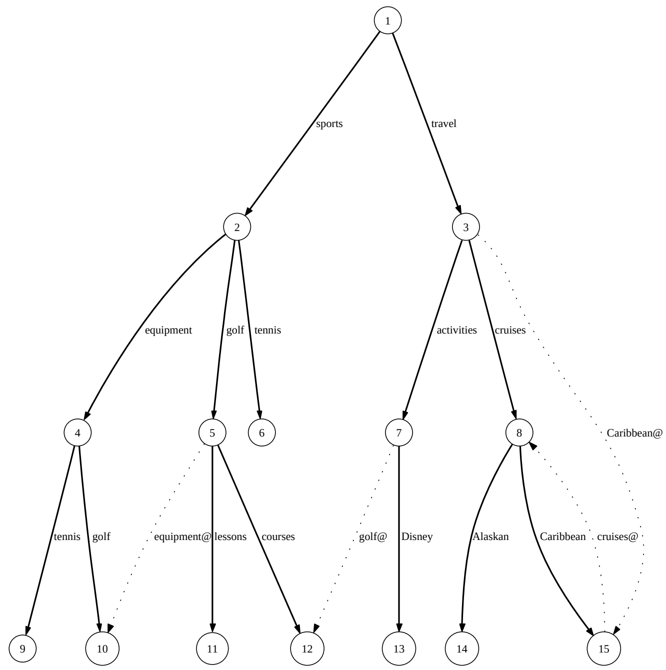

Figure 2: Example semistructured web directory sharing characteristics with *Yahoo!*. Symbolic links are indicated by dashed edges and hyperlink labels ending with @. All other typographical conventions follow those used in Fig. 1.

the total number of access paths to each terminal object in these sites is $\prod_{i=1}^{|f|} |v_i|$.

Often web directories, such as a *Yahoo!* and the *Open Directory Project* (*ODP*) at `dmoz.org`, are not faceted since, in these sites, no facet of information assessment classifies all of the edges at any given level of nesting. Fig. 2 illustrates such a site. Notice that unlike a faceted site, in this model, the link labels residing at the same depth are not related semantically. For instance, the link labels 'Disney' and 'lessons' at the third level of the site are not related aside from sharing (syntactic) depth. We call such sites *unfaceted*. The reader will notice that in Fig. 2, unlike Fig. 1, all paths through the site do not have a consistent length. For example, the $\prec$sports, tennis$\succ$ path has a length of 2 while all others have a length of 3. We call unfaceted sites with such a property *semistructured* for their similarities to models for semistructured data (Abiteboul et al., 2000). Web directories, which index sites across

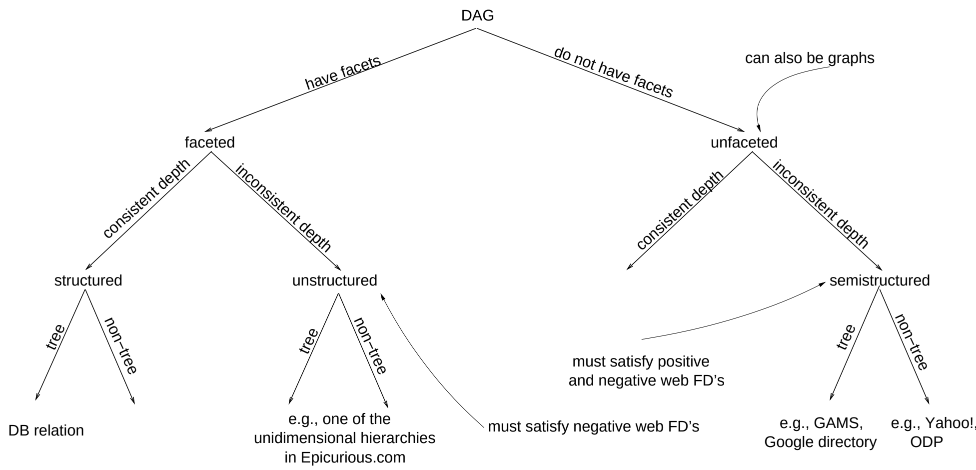

Figure 3: Classification of simple hierarchies described in this section.

the web on a variety of topics, use this underlying model. The classification used large web directories tends to be schemaless, self-describing, and semistructured. By virtue of being unfaceted, semistructured sites always satisfy (both positive and negative) web FD's. Fig. 3 graphically depicts these classes of information hierarchies with relevant annotations regarding distinctions. The *Google Directory* at `directory.google.com` has an underlying tree, semistructured data model. At the time of this writing, the author is unaware of any unfaceted sites with a consistent depth.

The primary purpose of web directories, such as *Yahoo!*, is to help users find a particular website. Broder (2002) has revealed that less than 50% of all web searches are informational in nature and has identified two additional types of web searches: navigational (trying to find a particular URL) and transactional. Through user and log analysis, Broder (2002) discovered that navigational search constitutes about 20% of searches. Web directories serve these navigational types of searches.

While information hierarchies provide a familiar interaction motif for users, they also are fixed and, thus, one-size-fits-all. If the hardwired classification of the items in the hierarchy (e.g., golf courses are classified in the sports category) does not match the user's mental model of the information space (e.g., 'I think of golf courses classified in the travel category), problems arise:

> "[F]inding the appropriate place within a category hierarchy can be a time-consuming task. [T]he user must look through a list of category labels and guess which one is most likely to

> contain references to the topic of interest. A wrong path requires backing up and trying again, and remembering which pages contain which information. If the desired information is deep in the hierarchy, or not available at all, this can be a time-consuming and frustrating process. Because documents are conceptually stored 'inside' categories, users cannot create queries based on combinations of categories using this interface" (Hearst, 1999).

One might be tempted to surmise that such problems only exist in semistructured sites, such as that shown in Fig. 2, because structured sites are more intuitive. Certainly, if a structured site was void of positive web FD's or, in other words, if all permutations of facet values were valid paths through the site, this problem would not exist (see Section 2). The user would only need to make an individual selection for each of the facets. However, in the presence of positive web FD's in structured sites, this underlying problem (i.e., the mismatch between the site's organization and the user's conception of the task) is the same as in semistructured domains, but manifests itself differently (e.g., 'I prefer to shop for computers using a *by display size, computer type, amount of memory* motif because I want to see if both desktops and laptops are available with a 15" display, but the site is organized in a *by computer type, display size, amount of memory* fashion and, therefore, I need to drill-down to desktops, check if a 15" screen is available, roll-up to the root, and follow the laptop link there to check for a 15" display').

We survey the three predominate, interrelated approaches to this problem — *faceted classification* (Section 2), a natural extension to it called *faceted search* (Section 3), and web directories with *symbolic links* (Section 4) — as well as tradeoffs between these approaches (see Table 2 and Fig. 14) and compositions of them (see Fig. 15). Faceted classification, faceted search, and symbolic links involve several aspects such as the underlying data model, operations on the model, and HCI, implementation (e.g., database, XSLT, and semantic web technologies), taxonomy design, and classification issues. However, this survey focuses only on the problem of supporting multiple access paths to each terminal object. Again, rather than being exhaustive in our presentation, we focus on broad, thematic issues.

2. Faceted Classification

An approach to accommodating multiple access paths to terminal objects in a web hierarchy is to enumerate all possible paths through the organization in the presentation of the hardwired hyperlink structure. Fig. 4 illustrates a rendering of the hierarchical model of Fig. 1 as a *faceted classification* (Wynar, 2000) with *hierarchical faceted categories* (Hearst, 2006a).

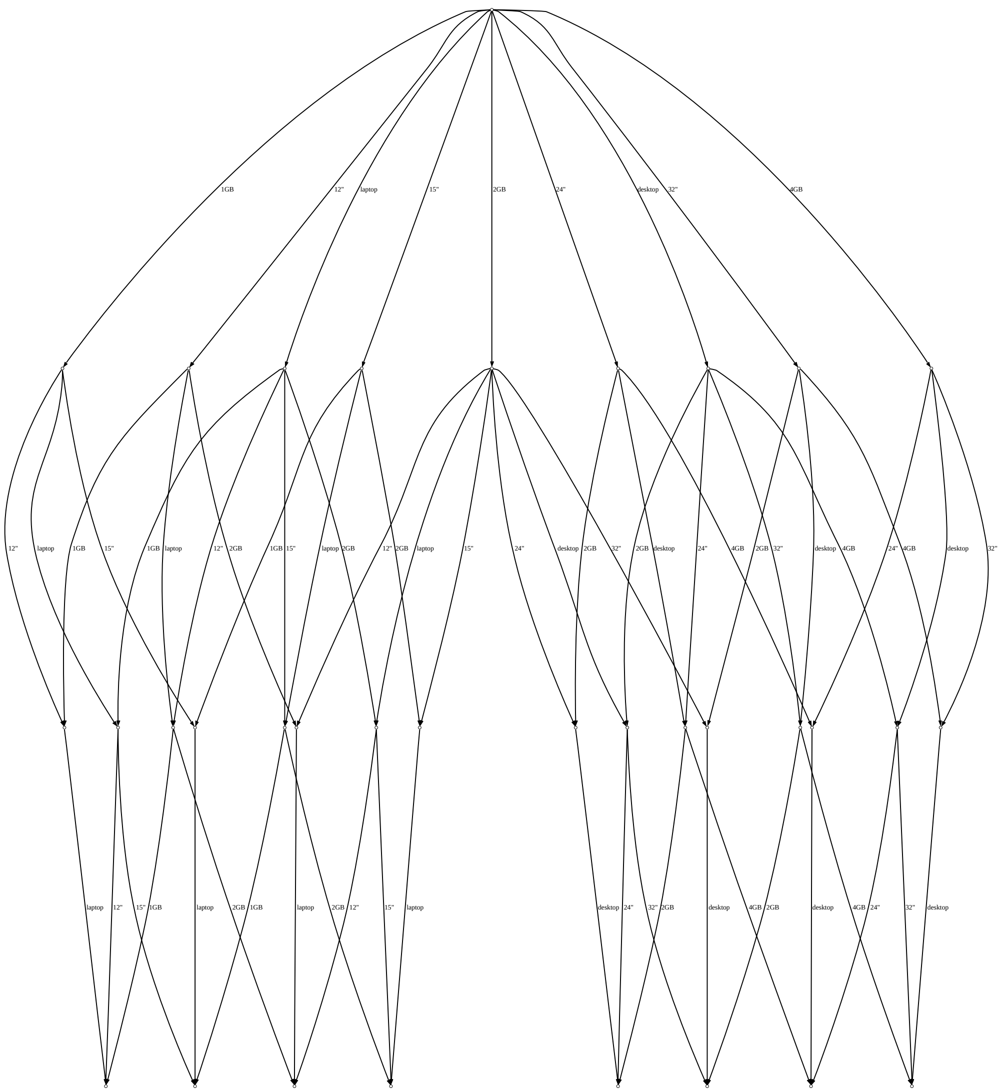

Figure 4: A sample faceted classification with characteristics similar to those in Epicurious.com. Uses the same typographical conventions as the DAG in Fig. 1.

A user browsing a site based on such a model is engaged in *faceted browsing*. 'A faceted classification differs from a traditional one in that it does not assign fixed slots to subjects in sequence, but uses clearly defined, mutually exclusive, and collectively exhaustive aspects, properties, or characteristics of a class or specific subject' (Wynar, 2000) called facets.

At this point, a note about terminology is required. We refer to faceted hierarchies such as that shown in Fig. 1 as *unidimensional* since such hierarchies only model one path to each terminal information object. Ranganathan (1951) refers to such organizations as *enumerative classifications*. In that context, the word *enumerative* refers to the enumeration of one presentation order of the facets along the paths of the hierarchy. On the other hand, we prefer to describe faceted classifications as enumerative because they enumerate all possible paths to each terminal object by permuting the presentation of the hyperlinks along the paths of the hierarchy. In other words, faceted classifications enumerate all possible unidimensional hierarchies in the presentation of the hyperlink structure and, therefore, are *multidimensional*. Such multiple intersecting hierarchies are referred to as *polyarchies* (Robertson et al., 2002) in the interactive information visualization community. Thus, in the context of hierarchy, there are two interpretations of the word *enumerative*.

We also prefer to avoid the use of the words *fixed* or *pre-determined* (Merholz, 2001) in discussing unidimensional hierarchies because, unless adapted dynamically during interaction (Brusilovsky, 2001; Perkowitz and Etzioni, 2000), any hypermedia hierarchy, including a faceted classification, is fixed or pre-determined. A unidimensional hierarchy can be faceted or unfaceted since unidimensionality is a syntactic property of the graph structure (i.e., a tree) while being faceted depends on how hyperlink labels at each level of the hierarchy are semantically related to each other. The advantages of faceted classifications over unidimensional organizations are discussed in (Aluri et al., 1991).

Interaction with systems and interfaces which support faceted classification can be thought of as user-directed while interaction with unidimensional hierarchies can be viewed as system-directed. Combining the two, as described below in Section 3, can lead to a mixed-initiative mode of interaction (Perugini and Ramakrishnan, 2003). 'Instead of sifting through a pre-determined hierarchy, the items are organized on-the-fly, based on their inherent qualities' (Merholz, 2001).

In contrast to a unidimensional classification, in a faceted classification, hyperlinks for all values of all facets with unspecified choices are available at each level of the site. For instance, at the root level of Fig. 4, hyperlinks for all choices of computer type, display size, and amount of memory are presented. Once a hyperlink at any level is followed, hyperlinks representing facet values for all facets filled at all previous levels are no longer available,[1] and so on until the user arrives at a leaf page. The implicit assumption is that the set of facet values selected, which can grow and shrink dynamically as the user interacts, form a conjunction of terms, and this interpretation of terms is the de facto standard in most faceted systems. Notice that the terminal objects, in this case, the computers, are not duplicated at the leaf level. Rather, each leaf page representing a computer is the target page of more than one edge. We shall return this point when we discuss symbolic links in Section 4.

Faceted classification has been used to organize items from reusable software components (Prieto-Díaz, 1991) to recipes at epicurious.com (see Fig. 5). Epicurious enumerates, in the presentation of its hardwired hyperlink structure, all possible permutations of each path to each recipe it indexes (akin to the sample classification shown in Fig. 4). For instance, at the root level of the classification (see Fig. 5, top), the user can make a selection from any of the facets (e.g., *recipe category*, *dietary consideration*, *cuisine*, and so on). At every point in the interaction, the user has the option to similarly refine by any of the facets remaining unfilled. For instance, in Fig. 5 (bottom), the user has the option to select a value for any of the *main ingredient*, *meal/course*, *type of dish*, *season/occasion*, *recipe category*, *preparation method*, and *dietary consideration* facets with an unspecified value. Notice hyperlinks representing values for the *cuisine* facet are unavailable in Fig. 5 (bottom) since the user made a selection (Thai) for the *cuisine* facet in Fig. 5 (top). In this manner, Epicurious accommodates multiple mental models of information-seeking or, in other words, allows users to experience any permutation of (the hyperlink labels in) any path of any corresponding unidimensional, faceted hierarchy of the recipes.

Fig. 6 illustrates the only two possible path permutations en route to the 'Mini Chocolate Cupcakes' recipe through the selection of a *type of dish*

---

[1] Some systems permit the user to select more than one value for the same facet, e.g., see `govconnection.com` and, therefore, in such cases, the facet values for the facet for which a value has been selected does not disappear. This approach can be considered an OR refinement, e.g., 'laptop OR desktop' for computer type.

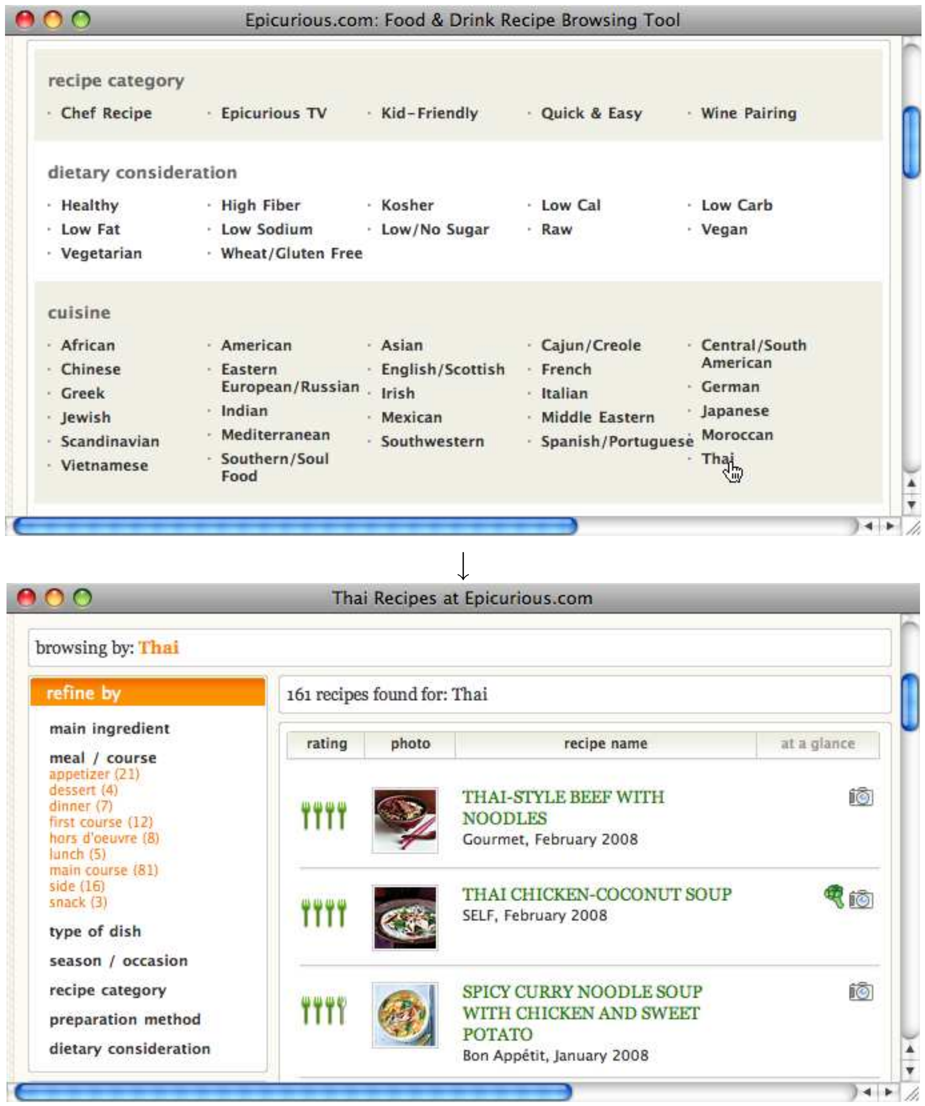

Figure 5: Faceted classification of recipes at Epicurious.com. The user has the option to choose a value for any facet remaining unfilled at any point in the interaction.

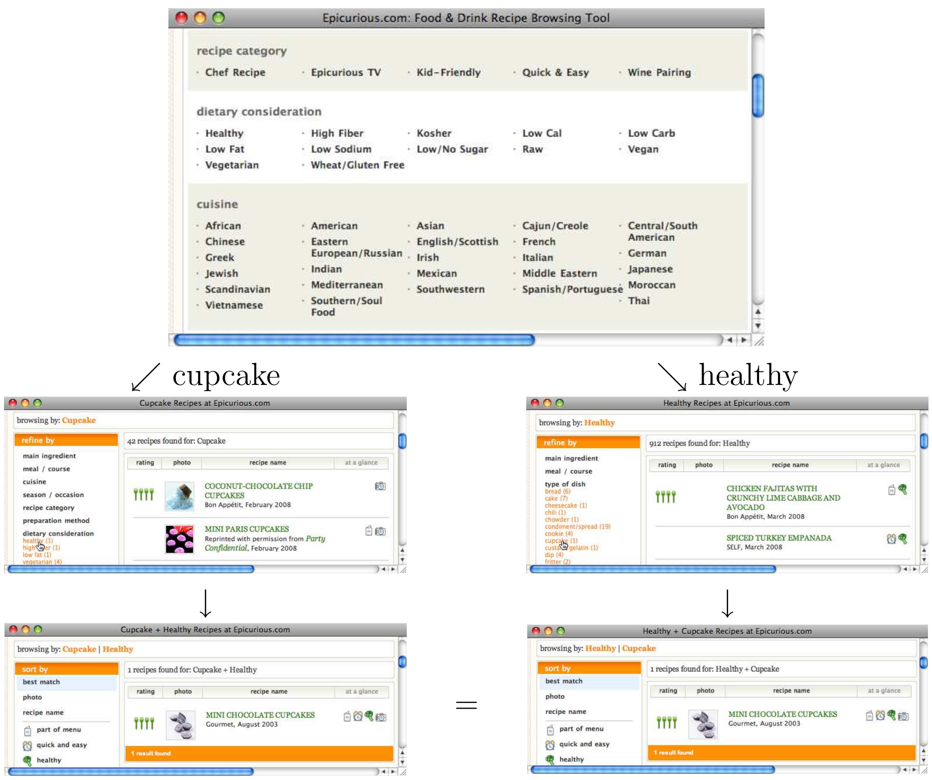

Figure 6: Path permutations in Epicurious.com. While both paths start at the same root (top) and end at the same recipe, the page on the bottom left is retrieved by following the ≺cupcake, healthy≻ path and the page on the bottom right is accessed through the ≺healthy, cupcake≻ path.

(cupcakes) and *dietary consideration* (healthy), i.e., $\prec$cupcakes, healthy$\succ$ (left) and $\prec$healthy, cupcakes$\succ$ (right), where a *path permutation* is a path whose ordered hyperlinks labels from root to leaf are a permutation of the ordered hyperlink labels of another path. There are a total of 8 facets in the Epicurious recipe classification. However, not all paths through Epicurious involve all 8 facets (e.g., the two leading to the 'Mini Chocolate Cupcakes' recipe). In other words, Epicurious does not index each recipe using a consistent number of facets, i.e., each of the multiple unidimensional hierarchies embedded into the classification is unstructured.

In Fig. 6, the series of links, i.e., 'Cupcake | Healthy' (bottom left) and 'Healthy | Cupcake' (bottom right), at the top of each page is called a *breadcrumb* and is used to provide the user with both context on where they have been and where they currently are (recall, one of the main complaints against hypertext is that users feel 'lost in hyperspace'; Nielsen, 1990), and thus reduce cognitive overhead, as well as easy access back to any previously visited pages. As described in Section 4, some links in a breadcrumb might provide access to pages not previously visited. Breadcrumbs are used to help orient users in hierarchies.

The number of access paths to each terminal object in a faceted classification of any faceted site F is described by the formula $\sum_{i=1}^{p} p_i!$, where p is the number of paths in F and p_i represents the number of facets (or, in other words, edges) involved in path i (Perugini, 2004). In a structured site, we can specialize this formula to $(\prod_{i=1}^{|f|} |v_i|) \times |f|!$, where f is the set of all facets and v_i is the set of facet values for facet f_i, to capture the number of access paths through the corresponding faceted classification. Notice that a faceted classification provides an organization with $O(|f|!)$ more access paths to each leaf than the analogous unidimensional, faceted site. Therefore, this approach involves a combinational explosion in the number of paths through the presentation of the classification as the number of facets, and facet values for each individual facet, increases. Thus, this approach is not practical, or perhaps even feasible, for devices with limited screen real estate (e.g., cell phones or PDA's). For approaches to mitigating this challenge in mobile devices see (Cohen et al., 2002; Dachselt and Frisch, 2007; Dakka et al., 2005; Karlson et al., 2006).

Not all information hierarchies can be rendered as a faceted classification. Since the concept of a facet does not exist in semistructured domains, a semistructured site cannot be rendered as a faceted classification. However, as we demonstrate in Section 4, symbolic links can help realize multiple access paths to terminal objects in such hierarchical sites.

The dependence of successive levels of a hierarchy on each other (Ben-Yitzhak et al., 2008) is a factor in the applicability of faceted classification. Specifically, some faceted sites, particular those modeled as IS-A hierarchies, cannot be rendered as faceted classifications. In general, as the number of positive web FD's satisfied by a hierarchy increase, the ease with which that hierarchy can be cast as a faceted classification decreases. IS-A hierarchies satisfy many positive web FD's. Consider an automobile IS-A hierarchy organized by *make* at the root-level and *model* at the following level. Since any value for the model facet determines the corresponding value for the make facet, web FD's such as 'Civic $\rightarrow$ Honda' are satisfied by the hierarchy (Perugini and Ramakrishnan, 2007). Therefore, it is unreasonable to solicit for make after model. Techniques for automatic query expansion (Perugini, 2006; Perugini and Ramakrishnan, 2007; White and Marchionini, 2006) are applicable in such cases. For more information, see the *stratum* metric in (Botafogo et al., 1992).

Some information hierarchies have levels which are neither purely IS-A nor completely independent. For instance, the wine shop at `wine.com`, a site which provides information on various types of wines, mixes faceted and IS-A classifications. At the root level one may browse by price, type, region, or professional ratings. However, at the root level, one may not browse by locality of a specific region because, e.g., Victoria IS-A region in Australia. In such spaces faceted classification can be used where appropriate.

Ultimately, in the absence of complete knowledge of users, a hierarchy designer attempting to accommodate a large segment of users must anticipate which of the $|f|!$ unidimensional hierarchies to enumerate in the presentation of hyperlinks. By supporting a faceted classification the designer is obviating the need to anticipate by expecting that users will require all possible unidimensional hierarchies. To avoid the mass hyperlink enumeration entailed in the presentation of a faceted classification, a designer might only selectively enumerate a subset of the $|f|!$ possible hierarchies (e.g., 'Click here to browse by computer type, display size, and amount of memory, or here to explore by display size, amount of memory, computer type'). Such an interface is however insufficient to support the information-seeking mental models of all users (e.g., those who prefer to browse by computer type, amount of memory, display size). To ensure that all users are accommodated we must enumerate (in the presentation of hyperlinks) all $|f|!$ browsing interfaces (e.g., 'Click here to browse by . . . , here to explore by . . . , . . . and here to browse by . . . .'). Again, as the number of facets increases, the required number of browsing hierarchies combinatorially explodes.

The are multiple projects and systems employing faceted classification. The *Flamenco* (FLexible information Access using MEtadata in Novel COmbinations) *Search System Project* (`http://bailando.sims.berkeley.edu/flamenco.html`) at UC Berkeley explores faceted classification in various catalogs and websites. Stoica et al. (2007) contribute the *Castanet* algorithm which can semi-automatically generate a faceted classification from textual descriptions of items. One can construct a facet map at `http://facetmap.com:8080/demo/index.jsp` to illustrate the advantages to faceted classification.

As we conclude this section, we feel it is helpful to provide some historical context from the field of information architecture. Many of the ideas discussed in this section and the prior section were developed in the first half the 20th century by the librarian S.R. Ranganathan whose primary contribution to classification theory most relevant to the discussion here is the *Colon Classification* system. It was through the Colon Classification that Ranganathan introduced the concepts of facets and facet analysis. Facets were Ranganathan's solution to the inherent problems with the *Dewey Decimal* and *Library of Congress* classification systems, namely, that index terms had to be conceived before an information object could be classified by the system. Rather than starting by designing a classification scheme and then fitting objects to it, Ranganathan's idea was to first start with the objects and then dynamically cull the index terms and attributes from them to provide more flexibility and description accuracy in the classification. The core facets which Ranganathan believed could classify any object were: personality, matter, energy, space, and time (Taylor, 1999). Facets values were separated by colons thereby giving the name to the system.

The rapid increase in the creation and subsequent use of massive amounts of digital information delivered through World Wide Web in the past fifteen years has provided an opportunity for the revitalization of Ranganathan's original classification ideas and caused a *classification renaissance*. Now faceted classification is the de facto standard for e-commerce websites and, particularly, online shopping catalogs, e.g., eBay and Amazon. There are both books (Morville and Rosenfeld, 2006) and articles (Kwasnick, 1999; Spiteri, 1998) discussing the application of fundamental information architecture ideas (classification theory and knowledge organization) to the web. Moreover, there is an XML markup language called XFML (Faceted Metadata Language) for building web faceted classifications. While the focus of this article is not on implementation issues, we refer the reader to (Denton, 2009) for an introduction on designing faceted classifications for use on the web.

3. Faceted Search

An alternate approach to support multiple access paths to each terminal object in an information hierarchy is *faceted search* (Hearst, 2006b). Faceted search is a natural extension to faceted classification. The fundamental idea in faceted search is to solicit and capture keywords supplied by a user from which to prune out branches of the hierarchy irrelevant to the user's information need. This style of search can be applied to both faceted (e.g., a unidimensional version of Epicurious) and unfaceted sites (e.g., ODP). Faceted search over a faceted site typically involves matching the terms in the query to the available values for the facets remaining unfilled to simplify the hierarchy at any point. Faceted search techniques integrate *navigational* (e.g., *Yahoo!*) and *direct search* (e.g., Google) to help users determine which portions of a classification contain the information desired. In other words, they combine browsing and search, leading to a mixed-initiative mode of interaction (Perugini and Ramakrishnan, 2003). While faceted classifications enumerate all possible path permutations to each terminal object in the presentation of hyperlinks, faceted search does not enumerate any, but rather uses search to achieve multiple access paths to the same item. While complementary in this respect, both approaches can be used in concert (leading to what is called *faceted browsing and search*) and, if implemented properly, the two integrate seamlessly. Faceted search is broad and refers to a family of related search techniques for information hierarchies as variations on this idea. In this section we discuss the idea and showcase a few specific examples.

Perhaps the most familiar instance of a faceted search interface is that used within Apple's popular *iTunes* digital audio software. The main window of the interface displays a relation over several music attributes (or facets) such as song name, artist, and album. This interface is not quite a pure faceted classification because making a selection within a facet in the browser does not prune the values for the facets presented to the left of the selection. Therefore, such interfaces are typically referred to as *directional faceted browsers* (Wilson et al., 2008) in that associations and pruning happen in a left-to-right fashion. However, the faceted search interface permits

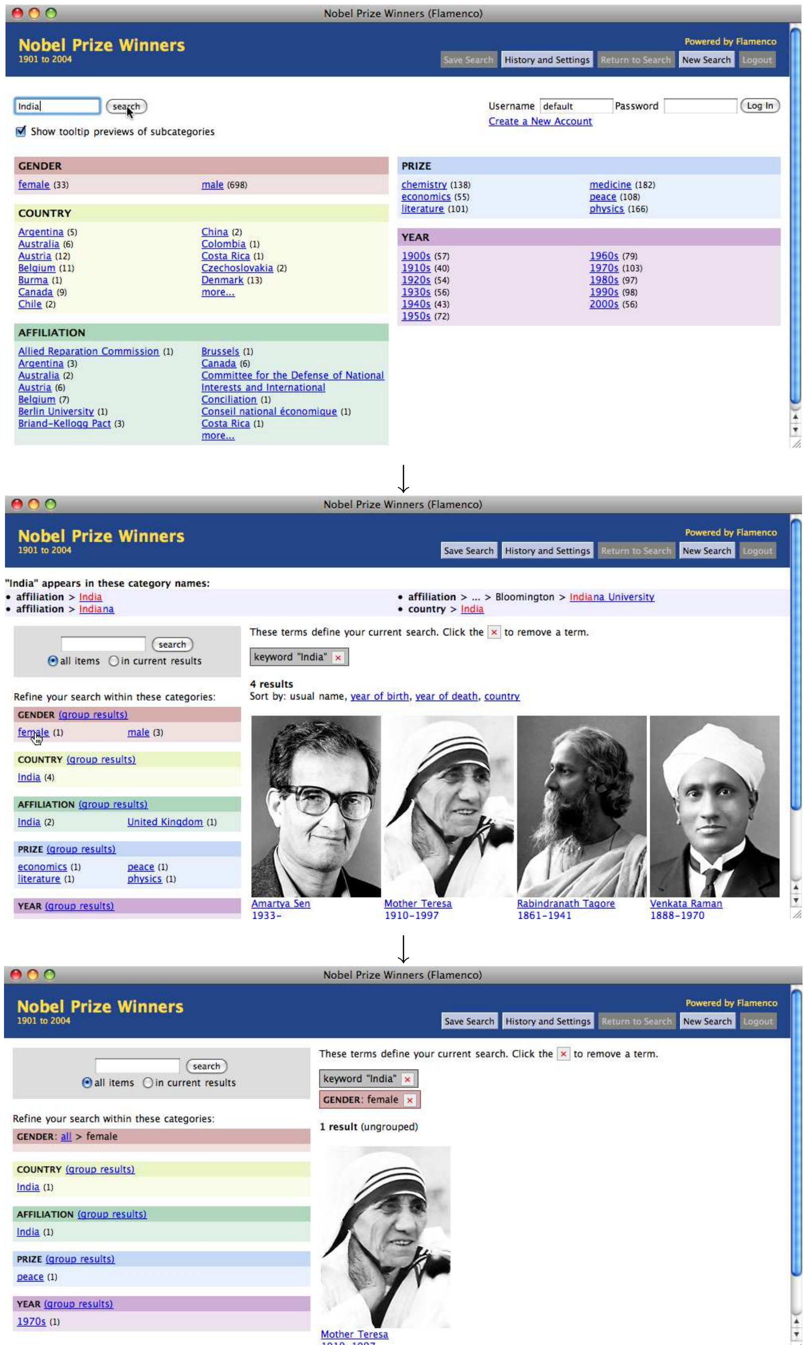

Figure 7: A sequence of interactions using the *Flamenco* faceted browsing and search interface.

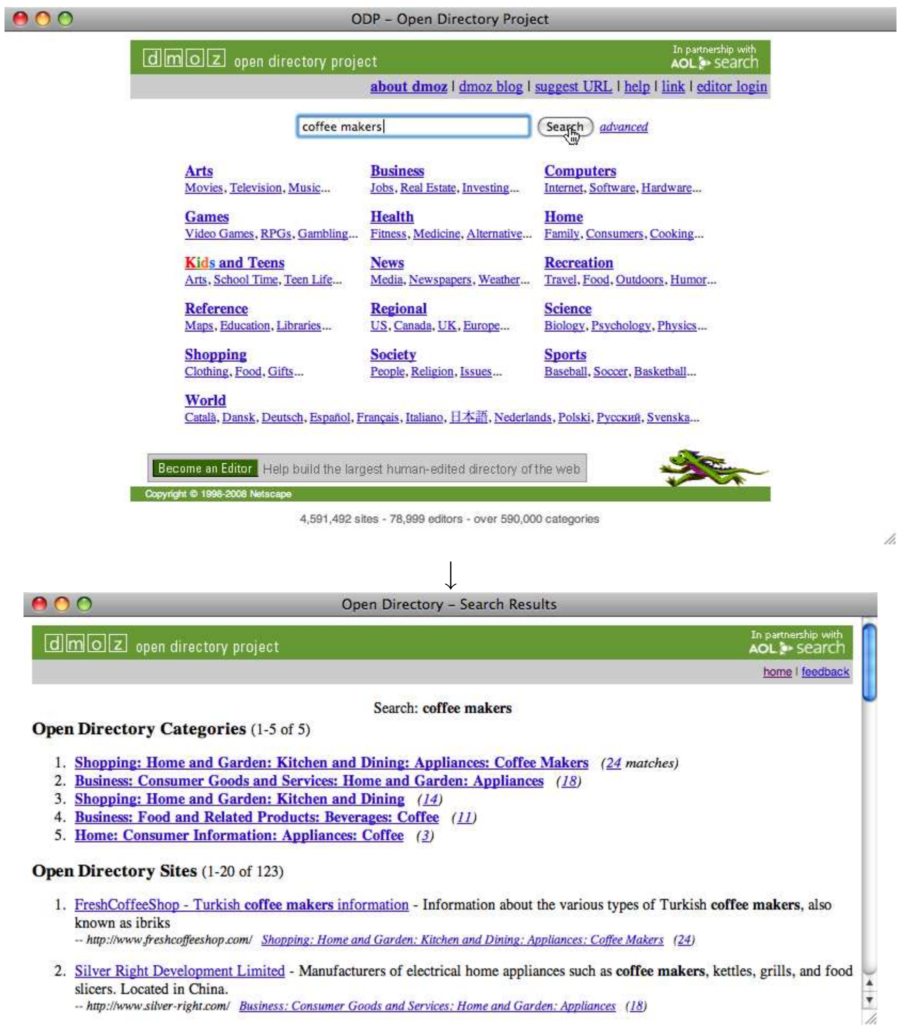

Figure 8: A search for ‘coffee makers’ using the ODP search facility. (top) Root page of the ODP directory with a query in the search textfield. (bottom) Results of the search.

the user to explore and prune the music library based on any or all of the available facets. The *Mail* e-mail application packaged with Apple's Mac OS X as well as the *Mozilla Thunderbird* e-mail client provide a similar faceted search. The *Relational Browser++* (*RB++*), also known as *RAVE* (Relation Attribute Viewer) (Zhang and Marchionini, 2004), incorporates similar ideas and was originally developed for Federal Statistical agencies.

Many faceted search interfaces are often used over faceted classifications. For instance, consider a *Flamenco* faceted search interface over the *Nobel Prize* winners (from `nobelprize.org`) illustrated in Fig. 7. At the root page (top window), the user searches the classification based on the term 'India.' This causes the country facet to be filled (see center window). Next the user clicks on the hyperlink labeled 'female' (center window) which fills the gender facet. That browsing interaction causes the remainder of the unfilled facets (e.g., affiliation, prize, and year, among others), by functional dependency, to become filled and leads the user to the leaf page for Mother Teresa (bottom window). Epicurious provides users a similar faceted search interface.

If implemented properly faceted browsing and search should seamlessly integrate with each other. Specifically, the processing of a hyperlink click in a faceted classification can be handled as the analogous search with the corresponding term and, therefore, browsing and search should be handled uniformly by the underlying implementation engine. However, while faceted browsing and search complement each other in interaction style, each independently support the same set of paths. Therefore, we feel that faceted search is most appropriate over a unidimensional hierarchy (see discussion of out-of-turn interaction below), where it supports the realization of paths not already modeled in the hyperlink structure, rather than a faceted classification, in a mobile domain (e.g., cell phone, PDA). While it may not be possible to display a faceted classification on a mobile device, due to limited screen real estate, it may be possible to display a unidimensional classification, and support faceted search on it to achieve multiple access paths to terminal objects.

The *Open Directory Project* (*ODP*) employs a search technique over its semistructured data model. Fig. 8 (top) demonstrates a search for 'coffee makers' from the root page of ODP. The results (shown in bottom window of Fig. 8) are a flat list of both sites and the categories in which those sites are classified in the directory. We see that sites about coffee makers are classified within the 'Shopping,' 'Business,' and 'Home' categories. The ODP search facility flattens the hierarchy, i.e., the semblance of hierarchy is lost as a result of a search. Moreover, the combination of browsing and search is not as rich as possible. First, the user is not permitted to explore a hierarchical rendition of the results (see Fig. 10). Second, the user cannot search again immediately within these results. Rather the user must first select one of the categories before the user can initiate another search within that portion of the directory. *Yahoo!* uses a similar search technique over its directory.

Sacco (2000) contributes the *zoom* operator for multidimensional taxonomies, i.e., those where a terminal item can be reached through more than one hardwired path. Zoom leverages the multiclassification in directories to help a user identify items of interest by thinning the hierarchal information base. Similarly, out-of-turn interaction (Narayan et al., 2004) is a technique for interacting with hierarchical websites which prunes links in the underlying site which do not lead to the desired information while preserving the semblance of a hierarchy. Specifically, out-of-turn interaction permits the user to retrieve any leaf webpage by communicating one or more hyperlink labels, using a voice or textual modality, in any order and, thus, unlike faceted search over a faceted classification, supports the realization of path permutations even though the corresponding hardwired paths may not actually exist within the presentation of the hierarchy. Fig. 9 illustrates an out-of-turn query for 'coffee makers' in ODP. Notice that out-of-turn interaction returns a webpage with only the links labeled 'Shopping,' 'Business,' and 'Home' and the user has the option to drill-down or interact out-of-turn yet again to prune hierarchy further from there. Out-of-turn interaction does not re-organize the hierarchy over which it is applied, but rather simply prunes it.

Zoom and out-of-turn interaction are approaches to faceted search which, unlike the ODP and *Yahoo!* search facilities, do not flatten the pruned hierarchy into a list of categories and curb subsequent interaction, but rather preserve the hierarchical structure (see Fig. 10) and permit the user to interleave hyperlink clicks (browsing) and searches in any order they desire at any point in the interaction leading to a mixed-initiative mode of interaction (Perugini and Ramakrishnan, 2003). With out-of-turn interaction the user can only experience by browsing the paths through the site explicitly hardwired into the structure. Any permutation of any of these browsing paths must be realized through a combination of browsing and search. Therefore, the paths realizable through out-of-turn interaction are untraversable by browsing. The cumulative effect of out-of-turn interaction automatically turns a unidimensional hierarchy (i.e., one with only one path to each terminal object) into a virtual multidimensional hierarchy containing all permuta-

Figure 9: A search for 'coffee makers' in ODP using out-of-turn interaction. See Fig. 10 for the entire resulting hierarchy.

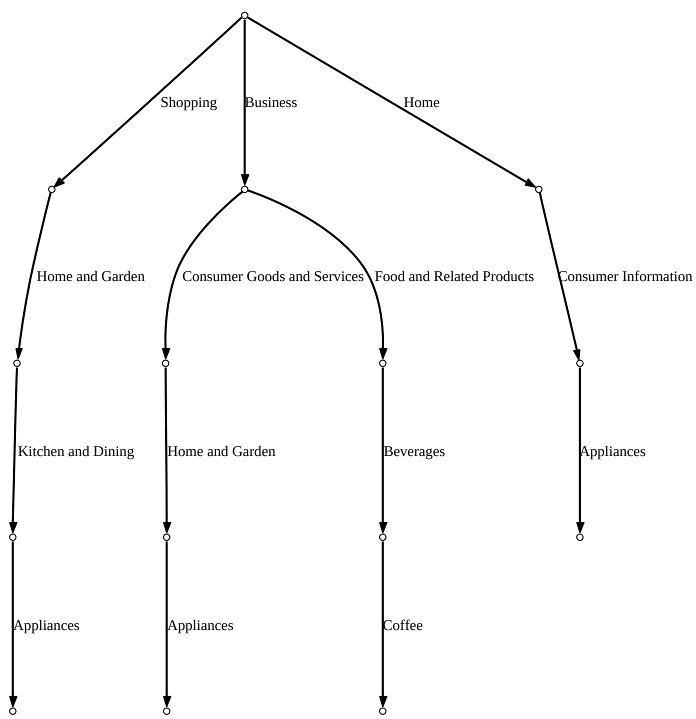

Figure 10: Schematic results, simplified for purposes of presentation, of a search for 'coffee makers' in ODP using out-of-turn interaction. Notice that the semblance of a hierarchy is preserved.

tions of (the hyperlink labels in) each path. We say virtual because the path permutations realized through out-of-turn interaction are available in effect (through search) and not in essence (through browsing).

A user can experience the exact same set of paths through both (i) the sample faceted classification given in Fig. 4 and (ii) applying out-of-turn interaction to the unidimensional, structured site in Fig. 1, albeit using different styles of interaction – purely browsing in the former and a combination of browsing and out-of-turn interactions (search) in the latter. Therefore, we can apply out-of-turn interaction over a structured incarnation of Epicurious with only one path to each recipe without compromising any of the paths within the faceted classification of Epicurious for the user. Since faceted classification involves a combinatorial explosion in the presentation of the hyperlink structure as the number of facets increases, faceted search over unidimensional hierarchies is particularly worthwhile for accessing systems with more than a few facets, especially in mobile environments where screen real estate is limited (Cohen et al., 2002; Dachselt and Frisch, 2007; Dakka et al., 2005; Karlson et al., 2006). Again, while these two approaches are functionally equivalent, each affords different interaction to the user. Therefore, which approach to employ in a solution depends on the goals of the system, targeted users, and empirical results from HCI studies.

Yee et al. (2003) offer information-seeking operators for faceted sites as part of the *Flamenco* system. *Magnet* (Sinha and Karger, 2005), which is part of the *Haystack* project at MIT, is a system for navigating semistructured information spaces. *ScentTrails* (Olston and Chi, 2003) is a project which seeks to combine browsing with search by adapting the presentation of links after an initial search, conducted to ascertain the information-seeking goal. Nevill-Manning et al. (1999) describe interfaces for navigating information hierarchies as well as techniques for automatically inferring them from large corpora. The common goal of these techniques and systems is to support flexible forms of interaction without requiring the designer to enumerate all possible modes of interaction into the presentation of the hardwired hyperlink structure as in faceted classifications.

While the faceted search approach avoids the enumeration involved in faceted classifications and, as a result, is applicable to devices with limited screen real estate, it has a disadvantage common to all search strategies: it requires the user to articulate their information need as set of terms, thereby compromising the principle advantage to hierarchy – familiarity with browsing. However, this problem is less severe in faceted than unfaceted domains because if users are aware of the facets, articulating an information need as a set of values for a subset of those facets is reasonable. In other words, in faceted search over faceted sites, users have the facets to guide query formulation.

Notice that permitting users to reach any terminal object by communicating facet values (i.e., hyperlink labels) in any order which corresponds to their conception of the task does not come for free. In a faceted classification, the designer must enumerate, in the presentation, hyperlinks corresponding to all possible permutations of the paths. In faceted search, a software module is required to support users in circumventing the hardwired organization of the original hierarchy en route to a particular terminal object.

Complete Answer Aggregates (Meuss and Schulz, 2001) are tree-based data structures used to facilitate browsing and querying tree-structured data. Hyperstructures (McGuffin and m.c. schraefel, 2004), such as *zzstructures*, *mSpaces*, and *polyarchies*, are data structures used for interacting with taxonomies which support operations such as similarity search.

While combining browsing and search (Manber et al., 1997) or, similarly, navigation and querying (Meuss and Schulz, 2001), is not a new concept, hierarchical websites, particularly in e-commerce domains, provide a practical and fertile landscape for their exploration. There have been three international workshops on the topic (the *2006 SIGIR Workshop on Faceted Search* and the *2007* and *2008 DEXA FIND* workshops). Moreover, there is a book (Sacco and Tzitzikas, 2009), an international community of researchers, and a special interest group (*SIG-FIND* at `http://www.sigfind.org`) dedicated to the topic.

4. Web Directory with Symbolic Links

Another approach to providing multiple access paths to terminal objects in information hierarchies involves *symbolic* (or soft) *links*. A symbolic link is a special hyperlink whose target vertex is the target of an existing hard link, where *hard links* create the natural parent–child relationships within a hierarchy. Symbolic links are sometimes referred to as cross-references in a hypertext context (Botafogo et al., 1992). Symbolic links are implemented in the UNIX filesystem, and a variety of other operating systems (e.g., MacOS, Windows), as a mechanism to create multiclassification without duplication (Marsden and Cairns, 2003). The file mode of a symbolic link in UNIX has a `l` in the leftmost bit, e.g., `lrwx------`. Symbolic links are used

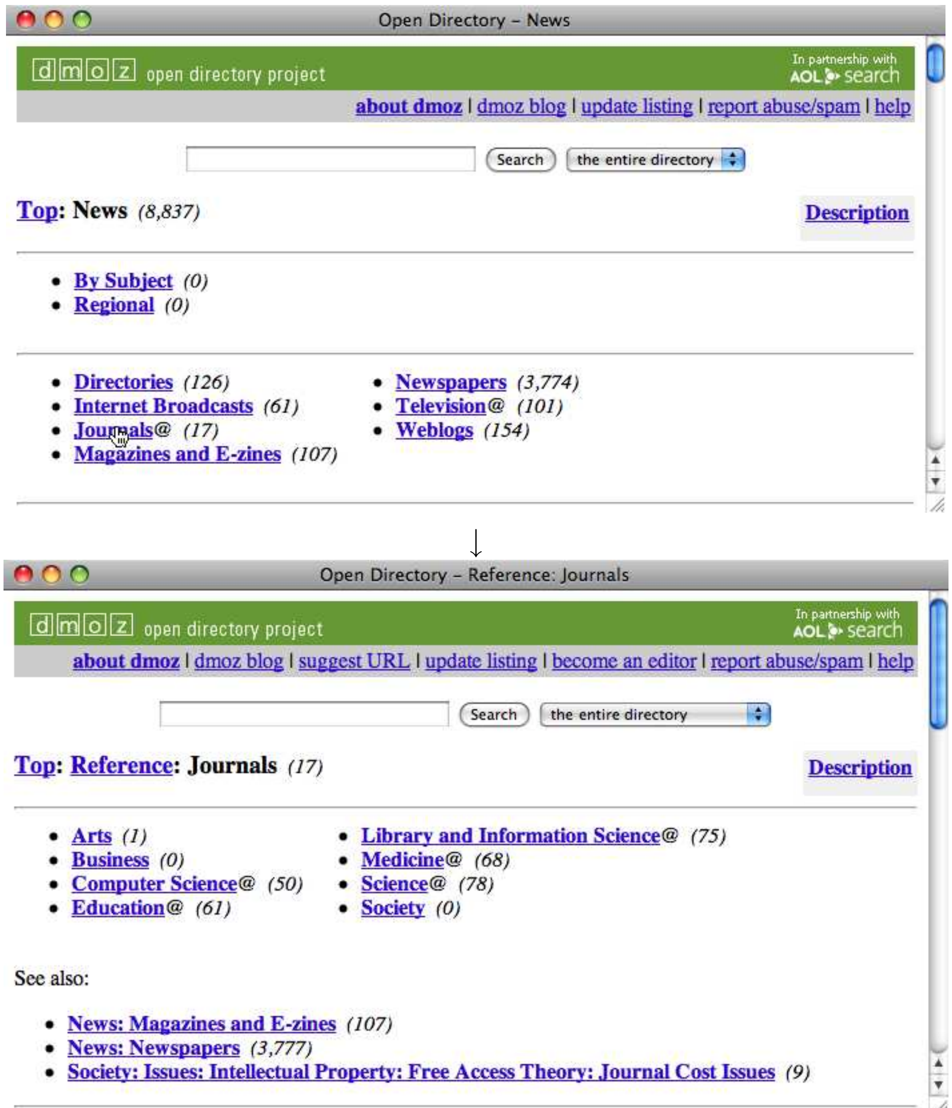

Figure 11: Illustration of a multiclassification symbolic link in the *Open Directory Project*. The user follows a symbolic link labeled 'Journals@' (top) which transports the user from the 'News' category (top) to the 'Reference: Journals' category (bottom).

predominately on the web in large directories such as *Yahoo!* and the *Open Directory Project* (*ODP*) at `dmoz.org` for similar purposes. The *Google Directory* at `directory.google.com` is an example of a web directory which does not employ symbolic links. The hyperlink label of a symbolic link in both *Yahoo!* and ODP ends with a @ character. The presence of symbolic links in web directories preclude such directories from being trees (Botafogo et al., 1992). In this section we identify 3 types of symbolic links, demonstrate how symbolic links are used in web directories to support multiple access paths to terminal objects, and compare symbolic links to other, similar hypermedia constructs.

4.1. Types of Symbolic links

There are three types of symbolic links in ODP: shortcuts, backlinks, and multiclassification links (Perugini, 2008). Re-consider the sample semistructured hierarchy in Fig. 2 which we use to describe these three types of symbolic links. Conceptually, a shortcut is an edge which provides direct access to a page nested at least one level deeper in a directory than its source. Formally, a navigational *shortcut* is a symbolic link whose target can be reached through a path, without symbolic links, through its source (Perugini, 2008). The edge from vertex 3 to 15 labeled 'Caribbean@' in Fig. 2 is a shortcut bypassing the intermediate page 8. A shortcut need not end in a leaf, i.e., a page in a directory with links only to external webpages (those which are not part of the directory and, therefore, are omitted from Figs. 2, 4, and 13 (to simplify presentation). Gerstel et al. (2007) contribute an algorithm for finding the set of *hotlinks* (i.e., shortcuts) which minimize the number of links a user must follow to access a leaf page in a web directory such as *Yahoo!* or ODP (called the *hotlink enhancement* problem). On the other hand, a backlink provides an avenue back to a more general category. For instance, in Fig. 2, the symbolic link labeled 'cruises@' is a backlink from webpage 15 to 8. Formally, a *backlink* is a symbolic link whose target can reach its source through a path without symbolic links (Perugini, 2008). The presence of backlinks in such directories preclude them from being DAG's due to the cycles that backlinks induce (see annotation of the *unfaceted* vertex label at the top right of Fig. 3). For instance, the reader will notice that the sample graph model of a semistructured site in Fig. 2 is not a DAG because of the cycle between page 8 and 15 induced by the backlink labeled 'cruises@.'

A multiclassification link is one which spans distinct categories in a directory. For instance, consider the symbolic link labeled 'golf@' from vertex 7 to

12 in Fig. 2. It creates an illusion that page 12 is categorized in the 'travel' section of the directory, when it is actually classified under 'sports.' 'The intended purpose this type of symbolic link is to tighten the gap between an item's actual placement in the directory and users' predictions of that placement during information seeking' (Perugini, 2008). A *multiclassification link* is formally a symbolic link whose target is classified under a different top-level category than its source. A *top-level category of category X* is any hyperlink label at the root page of category X (Perugini, 2008). For instance, 'sports' and 'travel' are top-level categories of the root in Fig. 2, 'golf' is a top-level category of the 'sports' sub-category (rooted at page 2), and 'activities' is a top-level category of the 'travel' sub-category (rooted at vertex 3). The symbolic link labeled 'equipment@' from page 5 to 10 is another multiclassification symbolic link in Fig. 2, but rather than bridging two top-level categories of the root, it spans two sub-categories of the 'sports' category ('equipment' and 'golf'). There are a total of 4 symbolic links in Fig. 2: 1 shortcut, 1 backlink, and 2 multiclassification links. Fig. 11 illustrates a multiclassification link in ODP, labeled 'Journals@,' which connects the 'News' category (top) to the 'Reference: Journals' category (bottom).

The reader will notice the 'see also' hyperlinks at the bottom of the webpage shown in Fig. 11 (bottom). These links, akin to symbolic links, are also cross-references to alternate sub-categories (in this case, to sub-categories in the 'News' and 'Society' categories). These 'see also' links indicate to the user, in a less transparent manner than symbolic links, that the 'Reference: Journals' category is related to the 'News: Magazines and E-zines,' 'News: Newspapers,' and 'Society: Issues: Intellectual Property: Free Access Theory: Journal Cost Issues' categories of the directory. In addition to ODP, 'see also' style cross-references are used in the *Guide to Available Mathematical Software* (*GAMS*) at `gams.nist.gov` – a semistructured web directory indexing mathematical software modules available in *Netlib* at *UTK* and *ORNL* at `netlib.org`.

4.2. Path Permutations

Perugini (2008) determined that symbolic links in ODP are used primarily for multiclassification. Table 1 provides a summary of the results from (Perugini, 2008). Specifically, 89% of the 97% of multiclassification symbolic links in ODP are within the same top-level category (see column labeled 'within'). In other words, a high frequency of the total number of symbolic links are within a top-level category. On the other hand, only 11%

	within	**between**
poorly connected	multiclassification links bridging top-level categories of top-level categories	
well connected	multiclassification links bridging topics sharing at least the first two levels of topic specificity	multiclassification links bridging top-level categories

Table 1: Summary of results of (Perugini, 2008).

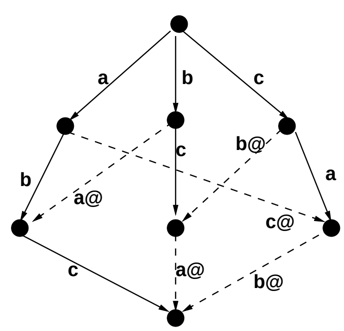

Figure 12: Sample (factored) faceted classification, simplified for purposes of presentation, illustrating permuted paths induced by symbolic links. Symbolic links help realize all 6 (=3!) permutations of the one and only ≺a, b, c≻ hard path (see leftmost path). Uses the same typographical conventions as Fig. 2.

of the multiclassification symbolic links are between top-level categories (see column labeled 'between'). This 11% of multiclassification symbolic links (which connect top-level categories) induce connections between 77% of all distinct top-level category connections possible (see 'well connected, between' cell). Moreover, the 89% of multiclassification symbolic links within the same top-level category only induce 17% of all possible distinct sub-categories of a particular top-level category (see 'poorly connected, within' cell). Lastly, approximately 90% of the 89% (or >77% of the total number of symbolic links) of multiclassification links within the same top-level category connect two categories which share at least the first two levels of topic specificity (see 'well connected, within' cell).

Symbolic links are kludges which deviate from the one-size-fits-all hierarchical classification induced by hard links in unidimensional organizations to help address situations where the rigid organization of the hierarchy is not appropriate for certain information-seeking tasks. Since all possible permutations of each path involving only hard links in a web directory are not present, such directories are not faceted classifications and, therefore, do not involve the hyperlink enumeration in the presentation of faceted classifications. However, multiclassification symbolic links can be used to induce the path permutations found in faceted classifications. For instance, consider page 10 in Fig. 2. It is reachable through the hard path $\prec$sports, equipment, golf$\succ$, where a *hard path* is a path without symbolic links. However, the symbolic link labeled 'equipment@' from vertex 5 to 10 helps realize an alternate view or permutation of (the link labels within) that hard path to page 10, i.e., $\prec$sports, golf, equipment$\succ$. Notice that the hyperlink labeled 'equipment@' is also a multiclassification link according to the definition given above. In other words, multiclassification links and symbolic links participating in path permutations are not mutually exclusive. All symbolic links participating in path permutations are multiclassification links, while the reverse does not hold. Symbolic links used in this manner help realize different permutations of one hard path and, therefore, help accommodate various models of information seeking users bring to bear upon their interaction with a directory. The onus is on the hierarchy designer to anticipate where a user might desire such a multiclassification symbolic link leading to a path permutation.

Alternatively, symbolic links can be used in this manner to create a factored faceted classification without compromising any path permutations. For instance, consider the sample faceted classification given in Fig. 12. Since there is only one hard path in Fig. 12, the use of symbolic links to foster permuted paths is more salient than in Fig. 4. Notice that in Section 2 we do not explicitly characterize the hyperlinks which induce permuted paths in Fig. 4 as symbolic links, i.e., they are not dotted and their labels do not end with a @ character. However, these links, whose target vertices have more than one parent, indeed help induce permuted paths. As of the time of this writing, the author is unaware of any faceted site or faceted classification which explicitly contains symbolic links in a visually identifiable way, i.e., with a @ character appended to the end of a hyperlink label. Using such implicit (or unidentifiable) symbolic links at the pre-leaf level to classify terminal objects through multiple paths is common in sites, which are otherwise trees, whose leaf vertices represent a flat list of (search) results (Perugini and Ramakrishnan, 2007). The fraction of symbolic links in directories such as ODP and *Yahoo!* which induce path permutations is currently unknown.

4.3. Yahoo!

Perhaps the most familiar example of a web directory with an underlying semistructured data model is *Yahoo!* (Glassel, 1998).

> "Each term in a *Yahoo!* notation string [referred to as a *breadcrumb* above] contains individual words which have meaning on their own, but once combined with other words into a string, a context is created, providing a deeper meaning. In this way it is much like a faceted classification.[2] One important advantage that virtual collections such as *Yahoo!* have over the print environment, in terms of notation schemes and their citation order (the order in which the facets are put together), is that the order of the facets in a string doesn't have to be set in stone. An electronic resource isn't limited to a single physical location. In a library, a book is only supposed to live in one place on a shelf. In the digital world, what is to stop us from classifying a resource in multiple places within a hierarchy? Nothing! By applying facet analysis to an online hierarchy, *Yahoo!* is able to take a string of categories and subcategories (facets) that describe a resource and by, rotating or [permuting] these terms, *Yahoo!* is able to provide access to a single resource from a variety of branches in the larger hierarchy.
>
> To illustrate *Yahoo!*'s ability to provide access to a single resource from two perspectives, I tried to locate information on the University of Wisconsin–Madison using the browsable hierarchy. I decided to start with the Regional category, thinking I would look for Wisconsin and then universities in Wisconsin. In the end, the category path I followed was:
>
> Regional : U.S. States : Wisconsin : Cities : Madison : Education :
> Colleges and Universities : University of Wisconsin – Madison

[2]Directories such as *Yahoo!* and ODP are often referred to as faceted sites or faceted classifications. We feel this characterization is misleading since the hyperlink labels at each level do not correspond to a facet of information categorization.

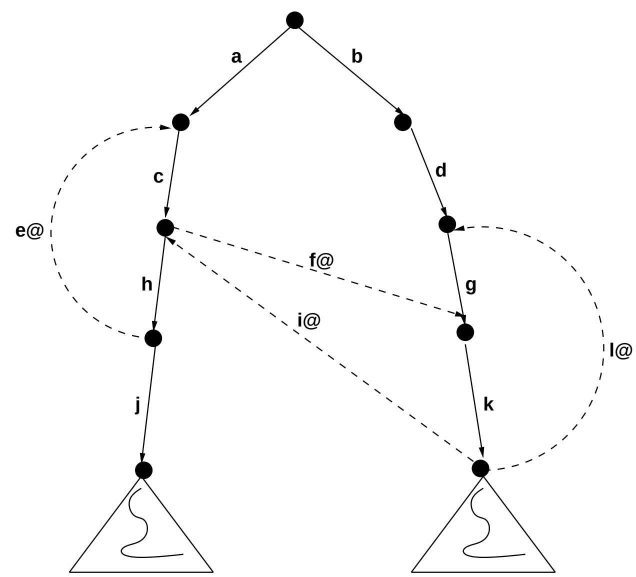

Figure 13: Sample semistructured web directory contrasting *backlinks* and *backjumps*. Uses the same typographical conventions as Fig. 2.

> I then wanted to see if I could find the same information by starting from the top of the hierarchy with the Education category. Once at 'Education: Higher Education: Colleges and Universities' the next link for the geographic subcategory ('United States@') actually drops me back into the Regional hierarchy, as noted initially. *Yahoo!* uses the @ symbol after any subcategory to indicate a cross reference, and also signifies that the link will take the user to another branch of the hierarchy in order to locate the information they seek. So in the end, anyone looking for the University of Wisconsin–Madison by entering the hierarchy by way of the Education category will be led to its home under the Regional category. But the important thing to remember is that no matter which path is followed initially, the user will end up with the information (Glassel, 1998)."

4.4. Related Link Types

There are multiple link types from the hypermedia literature related to symbolic links. For instance, we distinguish between *backjumps* from hypermedia and *backlinks* (also called *divergent* links; Rao and Turoff, 1990) – a type of symbolic link described above. A backjump is a link to the previously visited node (Bieber et al., 1997). Fig. 13 helps distinguish a backlink and backjump from each other and illustrates that multiclassification links can function as backjumps. Specifically, for the user who follows the ≺a, c, h, e@≻ sequence of hyperlinks, e is a backjump and backlink. Similarly, for the user who follows the ≺b, d, g, k, l@≻ sequence, l@ is a backjump and backlink. However, for the user who follows the ≺a, c, f, k, l@≻ sequence of hyperlinks, l@ is only a backlink. It is not a backjump because this user was never previously at the source of the link labeled g. Moreover, for the user who follows the sequence $\prec$a, c, f@, k, i@$\succ$, in addition to being a multiclassification link, i@ is a backjump.

A backjump should not be confused with the hyperlinks within a breadcrumb. In Fig. 11 notice the breadcrumbs 'Top: News' (top) and 'Top: Reference: Journals' (bottom) at the top of the pages. Breadcrumbs in unfaceted sites, as in this case, are as or more important to providing context than in faceted sites since unfaceted sites are void of facets to orient users. In the presence of symbolic links (as is the case in Fig. 11), if a user follows a multiclassification symbolic link in a browsing session, not all of the links in a subsequent breadcrumb may be backjumps, i.e., pointers back to previously visited pages, because symbolic links destroy the context built up into a breadcrumb and, therefore, actually exacerbate the 'lost in hyperspace' problem (Nielsen, 1990). For instance, consider a user who arrives at page 12 in Fig. 2 by following the (hard) path $\prec$travel, activities, golf$\succ$. If this user follows the 'golf' link in the breadcrumb at page 12, the user will be transferred to page 5 rather than page 7 because breadcrumbs only contain hard links. However, this user had not been at page 5 en route to page 12. Therefore, symbolic links in web directories compromise orientation and context. An accurate breadcrumb for this user at page 12 in Fig. 2 is 'Top: Travel: Activities: Golf.' However, since this breadcrumb does not correspond to an actual category in the directory, if used, the one-to-one correspondence between breadcrumbs and categories in the directory would be lost. ODP and *Yahoo!* suffer from this same problem. Notice that the breadcrumbs in Epicurious (see Fig. 6) preserve navigation history. This reinforces the fact that symbolic links are indistinguishable from hard links in Epicurious.

Composites and *transclusions* (or *inclusions*) are hypermedia constructs related to symbolic links. Both can be thought of as collection data structures (Helic et al., 1999) which permit containers (e.g., directories) to share the same components without duplication (Bieber et al., 1997). In this manner, like symbolic links, both provide a single access point for modifications obviating the need to synchronize updates across multiple copies of the same data (Perugini, 2008). However, composites, like symbolic links, only support the sharing of whole items. On the other hand, transclusions, unlike symbolic links and composites, can refer to subparts of data objects (e.g., paragraphs).

Warm and *hot* links, additional hypermedia constructs, are similar to

	faceted classification	faceted search	symbolic links
flexibility	$\surd$	$\surd$	$\surd$
disorientation	$\times$	some	$\surd$
context switch	$\times$	$\surd$	$\times$
enumeration	$\surd$	$\times$	some
anticipation	$\times$	$\times$	$\surd$
goal articulation	$\times$	$\surd$	$\times$

Table 2: A comparison of faceted classification, faceted search, and symbolic links across various attributes.

transclusions conceptually, but use actual copies of data rather than pointers and, therefore, are the antithesis of symbolic links. Users must explicitly request an update of the data when using warm links, while content is verified and automatically updated with hot links (Bieber et al., 1997). Often transclusions and warm/hot links are used in concert. Transclusions avoid duplication, while warm/hot links provide easy (local) access (Bieber et al., 1997). While transclusions and warm/hot links have not been embraced by hypermedia implementors (Bieber et al., 1997), composites have been introduced to the web (Helic et al., 1999). In addition, cross-references and composites have been used in hypermedia applications, including one built using web technologies (Balasubramanian et al., 1997).

Researchers have studied how to cast search results in the context of a web directory. For instance, Dumais and Chen (2000) used hierarchical classifications, with machine learning algorithms, to classify web search results. *DirectoryRank* (Krikos et al., 2005) is a system for ordering pages on a particular topic in ODP-like taxonomies. Other researchers have developed techniques to predict category accesses by a user in an information taxonomy (Chen et al., 2002).

This section has surveyed research related to symbolic links. In summary, symbolic links have their roots in hierarchical databases and filesystems where they are used for multiclassification without duplication, symbolic links have evolved to be used in hypermedia and web directories for similar purposes, research on hyperlinks in general is limited (nodes have been the primary focus), the little research on links has focused on semantic link typing, and we identified that multiclassification links can help realize permuted paths in web directories.

5. Discussion

5.1. Summary

We have presented the three predominate, interrelated approaches to supporting multiple access paths to each terminal object in information hierarchies. Table 2 compares faceted classification, faceted search, and web directories with symbolic links across a variety of advantages and disadvantages. While there are minor differences in the implementation and presentation details of each approach (e.g., out-of-turn interaction preserves the hierarchical metaphor while the ODP search does not), we offer Table 2 as a conceptual comparison. All three approaches provide support for flexible information access. Symbolic links often destroy context and thus their use in directories exacerbates the 'lost in hyperspace' problem (Nielsen, 1990). On the other hand, faceted classification does not suffer from problems of disorientation because of the structured and intuitive environment (supported by hypermedia constructs such as breadcrumbs) in which the facets are presented and pruned during browsing. However, disorientation is an issue with some faceted search techniques such as out-of-turn interaction. Nonetheless, faceted search over a unidimensional classification is an approach to dealing with the limited screen real estate of mobile computing devices.

Symbolic links integrate seamlessly with the familiar browsing paradigm, whereas faceted search requires a context switch between browsing and searching, and vice versa. Of course, faceted classification requires the designer to enumerate all $|f|!$ unidimensional organizations in hyperlink presentation. This enumeration can be reduced by using implicit symbolic links as demonstrated in Figs. 4 and 12. Faceted search is not plagued by this problem. However, including symbolic links in a web directory does involve some enumeration since the symbolic links must be explicitly included in the presentation of the hyperlink structure. Another way to view this issue is to say that web directory designers seeking to accommodate multiple user models of information seeking with symbolic links need to anticipate where (in the directory) the user might desire a symbolic link. Recall that designers of directories are faced with the challenge of organizing its structure so that users can find the information they seek in an natural, intuitive, effective, and efficient manner. Including symbolic links is an approach to this challenge, but forces the designer to manually anticipate the points at which the user might find a symbolic link helpful. Faceted classification obviates the need for such anticipation by enumerating all possible paths, in the presentation

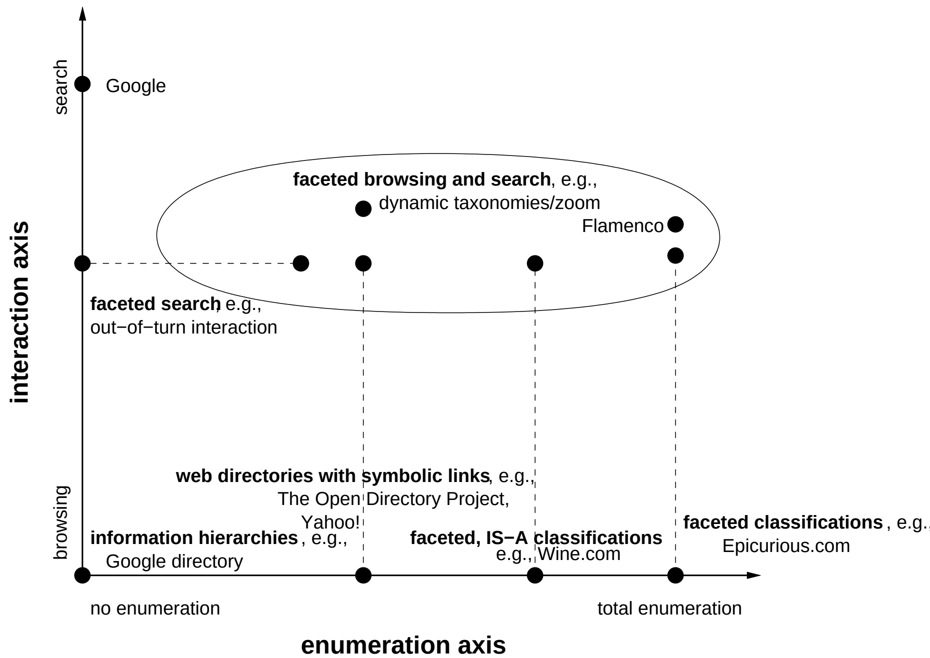

Figure 14: Graphical overview of tradeoffs.

of the hyperlink structure, to each terminal object and, therefore, achieves no anticipation by total enumeration. Faceted search also does not involve anticipation. However, use of a faceted search interface requires the user to articulate their information-seeking goal as a set of terms, which can be a challenge, but less so in faceted domains versus unfaceted environments. Faceted classification and symbolic links are browsing constructs and, therefore, do not require goal articulation as a set of terms. The graph in Fig. 14 provides an alternate view of the tradeoffs in these approaches across interaction and enumeration axes. The four points residing within the 'faceted browsing and search' region which are connected to dotted lines reinforce that a search facility can be added to web directories and faceted classifications.

Fig. 15 graphically summarizes the three approaches showcased in this article. The enumeration label on the lhs is intended to indicate complete enumeration in contrast to the (incomplete) sprinkling of symbolic links in web directories depicted on the rhs. These three approaches are not mutually exclusive. For instance, the ODP and *Yahoo!* directories both use symbolic links and a search facility, and Epicurious employees all three approaches.

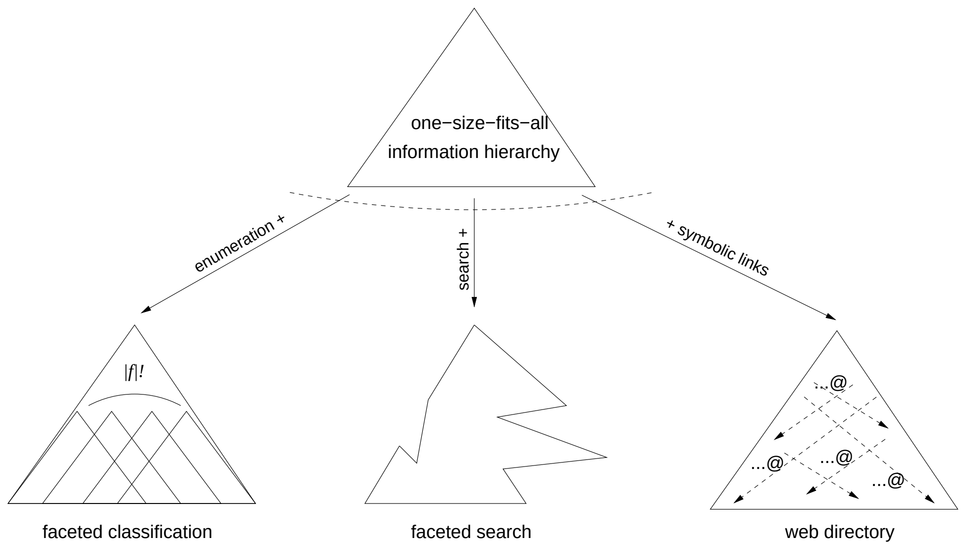

Figure 15: Approaches to supporting multiple access paths to each terminal object in information hierarchies. (left) information hierarchy + enumeration = faceted classification. (center) information hierarchy + search = faceted search. (right) information hierarchy + symbolic links = web directory.

For these reasons, we include a dotted arc in Fig. 15 spanning all three approaches to indicate combinations are possible.

5.2. Future Research Directions

The current thrust of research in faceted browsing and search focuses on user interface improvements and connections not only to other search techniques, but also to other area of computer science such as data mining. Some broad directions for future research include:

- **User Interface and Visualization Enhancements**: Novel user interfaces to augment standard faceted browsing and search interfaces with, for example, visualizations for faceted filtering (Weiland and Dachselt, 2008) and restrictions (Tvarozek and Bielikova, 2008), have been proposed and studied. Automatic construction of multi-faceted interfaces is also a direction with promise (Dakka et al., 2005).

- **Knowledge Discovery in Faceted Browsing and Search**: Recently, research projects aimed at revealing latent associations and trends in faceted and hierarchical data to support information-discovery tasks have emerged. Some projects propose to augment faceted interfaces with visualization and interaction facilities for knowledge discovery, e.g., *FacetLens* (Lee et al., 2009). The *backward highlighting* technique allows users to explore relationships in faceted data in a *iTunes*-like browser in a right-to-left manner in addition to the traditional left-to-right direction (Wilson et al., 2008). The *FacetPatch* browser explicitly recognizes that search is a procedural activity in which requirements evolve during user interaction. The *contextual facets* technique (Medynskiy et al., 2009), which is explored in *FacetPatch*, overlays related facets from external webpages into the current page and can be thought of as the analog of similarity search in faceted environments. Work has also been done on adding OLAP-type capabilities, beyond simple information previews such as item counts within each facet value, to faceted browsing and search systems (Ben-Yitzhak et al., 2008). Such research will become more important with the rise in popularity and use in recent years of social networking sites (e.g., Facebook, Twitter).

- **Integration with Alternate Search Paradigms**: Since search tasks are complex and often involve structuring sub-searches, it should be no surprise that faceted browsing and search is no silver bullet. Therefore, research is being done on how to best combine faceted search with other search paradigms such as full-text search (e.g., Google) and content-based query-by-example search (Tvarožek and Bieliková, 2008; Wilson and m.c. schraefel, 2008). Conversely, information taxonomies can also be used to improve full-text search (Fontoura et al., 2008).

- **Application Domains Beyond Websites and Digital Libraries**: While websites and digital libraries rich in meta and relational data provide a fertile landscape for the exploration of faceted browsing and search techniques, the area has expanded to geographic information systems (GIS) and returned to improving interaction with filesystems. Research is underway on extending *Flamenco* to include spatial, in addition to textual, functionality (Frontiera, 2008). Integrating faceted techniques into petabyte-scale filesystems is an approach to familiarizing oneself with the entire contents of a large disk (Koren et al., 2007). It is still an open research question as to how to naturally integrate faceted browsing and search interfaces and techniques into social networking systems where the semblance of an information hierarchy is absent.

This list is not exhaustive and the boundaries of these areas are becoming fuzzy with respect to both research and development. For example, the *FacetBrowser* project (Dachselt and Frisch, 2007) explores user-interface enhancements, serendipitous associations, and complex, procedural search tasks. Moreover, as described above, integration goes beyond the incorporation of other search paradigms and extends to seemingly orthogonal areas such as data mining and GIS.

5.3. Contributions

With the rise in recent years in the volume and concomitant use of websites presenting products and services, faceted browsing and search interfaces have become the de facto standard for e-commerce (one needs to look no further than *eBay*). We have provided a compass for both designers and users to navigate this space. Specifically, we explored and provided a high-level, accessible overview of the faceted browsing and search, and symbolic link, space, offered a new perspective by positing that web directories such as *Yahoo!* and ODP are not faceted sites, incorporated symbolic links into the discussion of faceted classification (the faceted browsing and search community has been largely silent on the issue of symbolic links), and cast symbolic links as hypermedia constructs which help induce path permutations which can, in turn, if used strategically, lead to a compacted faceted classification. The tradeoffs between and the novel possible compositions of the three approaches surveyed in this article lead to research questions:

- Should a designer create a faceted classification, i.e., enumerate all possible unidimensional hierarchies in the presentation of the hyperlink structure, or a partial faceted classification, i.e., only enumerate a subset of all possible unidimensional hierarchies?

- What is the most appropriate method to support all possible path permutations: (i) a fully expanded faceted classification, (ii) a faceted search technique, such as out-of-turn interaction, over a unidimensional hierarchy, or (iii) a faceted classification compacted with symbolic links?

- How are symbolic links most appropriately used: (i) as shortcuts, back-links, or multiclassification links in web directories, or (ii) to induced all possible permuted paths as in compacted faceted classifications? In other words, the tradeoff between symbolic links (partially enumerated) and faceted classifications is currently an open research issue. Where to stop is unknown.

Each of these research questions is multi-faceted and, thus, involves aspects such as system goals, targeted users, and empirical results from HCI studies.

Acknowledgments

This work was supported in part by seed grants from the University of Dayton Research Council, the Ohio Board of Regents, and the University of Dayton College of Arts and Sciences.

Vitae

Saverio Perugini is an Assistant Professor in the Department of Computer Science at the University of Dayton. His research interests include information personalization, web mining, and functional programming. He has a Ph.D. in Computer Science from Virginia Tech.

References

Abiteboul, S., Buneman, P., Suciu, D., 2000. Data on the Web: From Relations to Semistructured Data and XML. Morgan Kaufmann, San Francisco, CA.

Aluri, R., Kemp, D., Boll, J., 1991. Subject Analysis in Online Catalogs. Libraries Unlimited, Inc., Englewood, CO.

Baeza-Yates, R., Ribeiro-Neto, B., 1999. Modern Information Retrieval. Addison-Wesley, Boston, MA.

Balasubramanian, V., Bashian, A., Porcher, D., 1997. A Large-scale Hypermedia Application using Document Management and W eb Technologies. In: Bernstein, M., Carr, L., Østerbye, K. (Eds.), Proceedings of the Eighth Annual ACM Conference on Hypertext. ACM Press, New York, NY, pp. 134–145.

Ben-Yitzhak, O., Golbandi, N., Har'El, N., Lempel, R., Neumann, A., Ofek-Koifman, S., Sheinwald, D., Shekita, E., Sznajder, B., Yogev, S., 2008. Beyond Basic Faceted Search. In: Najork, M., Broder, A., Chakrabarti, S. (Eds.), Proceedings of the International Conference on Web Search and Web Data Mining (WSDM). ACM Press, New York, NY, pp. 33–44.

Bieber, M., Vitali, F., Ashman, H., Balasubramanian, V., Oinas-Kiukkonen, H., 1997. Fourth Generation Hypermedia: Some Missing Links for the World Wide Web. International Journal of Human-Computer Studies 47, 31–65.

Botafogo, R., Rivlin, E., Shneiderman, B., 1992. Structural Analysis of Hypertexts: Identifying Hierarchies and Useful Metrics. ACM Transactions on Information Systems 10 (2), 142–180.

Broder, A., 2002. A Taxonomy of Web Search. SIGIR Forum 36 (2), 3–10.

Brusilovsky, P., 2001. Adaptive Hypermedia. User Modeling and User-Adapted Interaction 11 (1–2), 87–110.

Chen, M., LaPaugh, A., Singh, J., 2002. Predicting Category Accesses for a User in a Structured Information Space. In: Belkin, N., Ingwersen, P., M.-K.Leong (Eds.), Proceedings of the Twenty-fifth Annual International

ACM Conference on Research and Development in Information Retrieval (SIGIR). ACM Press, New York, NY, pp. 65–72.

Cohen, D., Herscovici, M., Petruschka, Y., Maarek, Y., Soffer, A., 2002. Personalized Pocket Directories for Mobile Devices. In: Lassner, D. (Ed.), Proceedings of the Eleventh International World Wide Web Conference (WWW). ACM Press, New York, NY, pp. 627–638.

Dachselt, R., Frisch, M., 2007. Mambo: a Facet-based Zoomable Music Browser. In: Ojala, T. (Ed.), Proceedings of the Sixth International Conference on Mobile and Ubiquitous Multimedia (MUM). ACM Press, New York, NY, pp. 110–117.

Dakka, W., Ipeirotis, P., Wood, K., 2005. Automatic Construction of Multi-faceted Browsing Interfaces. In: Herzog, O., Schek, H.-J., Fuhr, N., Chowdhury, A., Teiken, W. (Eds.), Proceedings of the Fourteenth ACM International Conference on Information Knowledge Management (CIKM). ACM Press, New York, NY, pp. 768–775.

Denton, W., Jun. 17, 2009. How to Make a Faceted Classification and Put It On the Web. URL: `http://www.miskatonic.org/library/facet-web-howto.html`.

Dumais, S., Chen, H., 2000. Hierarchical Classification of Web Content. In: Belkin, N., Ingwersen, P., M.-K.Leong (Eds.), Proceedings of the Twenty-third Annual International ACM Conference on Research and Development in Information Retrieval (SIGIR). ACM Press, New York, NY, pp. 256–263.

Fontoura, M., Josifovski, V., Kumar, R., Olston, C., Tomkins, A., Vassilvitskii, S., 2008. Relaxation in Text Search using Taxonomies. Proceedings of the VLDB Endowment 1 (1), 672–683.

Frontiera, P., 2008. Flamenco + Geo: Extending a Hierarchical Faceted Metadata Search Interface with Geographic Capabilities. In: Aref, W., Mokbel, M., Schneider, M. (Eds.), Proceeding of the Second International Workshop on Geographic Information Retrieval (GIS). ACM Press, New York, NY, pp. 55–56.

Garcia-Molina, H., Ullman, J., Widom, J., 2001. Database Systems: The Complete Book. Prentice Hall, Upper Saddle River, NJ.

Gerstel, O., Kutten, S., Laber, E., Matichin, R., Peleg, D., Pessoa, A., Souza, C., 2007. Reducing Human Interactions in Web Directory Searches. ACM Transactions on Information Systems 25 (4), 1–27, article 20.

Glassel, A., Mar. 1998. Was Ranganathan a Yahoo!? InterNIC NewsIn End User's Corner column. URL: http://scout.cs.wisc.edu/addserv/toolkit/enduser/archive/1998/euc-9803.html.

Hearst, M., 1999. Modern Information Retrieval. Addison-Wesley, Boston, MA, Ch. 10: User Interfaces and Visualization, pp. 257–323.

Hearst, M., 2006a. Clustering versus Faceted Categories for Information Exploration. Communications of the ACM 49 (4), 59–61.

Hearst, M., 2006b. Design Recommendations for Hierarchical Faceted Search Interfaces. In: Broder, A., Maarek, Y. (Eds.), Proceedings of the International ACM SIGIR Workshop on Faceted Search.

Helic, D., Maurer, H., Scherbakov, N., 1999. Introducing Hypermedia Composites to WWW. Journal of Network and Computer Applications 22 (1), 19–32.

Karlson, A., Robertson, G., Robbins, D., Czerwinski, M., Smith, G., 2006. FaThumb: A Facet-based Interface for Mobile Search. In: Cockton, G., Korhonen, P. (Eds.), Proceedings of the ACM Conference on Human Factors in Computing Systems (CHI). ACM Press, New York, NY, pp. 711–720.

Koren, J., Leung, A., Zhang, Y., Maltzahn, C., Ames, S., Miller, E., 2007. Searching and Navigating Petabyte-Scale File Systems based on Facets. In: Gibson, G. (Ed.), Proceedings of the Second International Workshop on Petascale Data Storage (PDSW). ACM Press, New York, NY, pp. 21–25.

Krikos, V., Stamou, S., Kokosis, P., Ntoulas, A., Christodoulakis, D., 2005. DirectoryRank: Ordering Pages in Web Directories. In: Bonifati, A., Lee, D. (Eds.), Proceedings of the Seventh Annual ACM International Workshop on Web Information and Data Management (WIDM). ACM Press, New York, NY, pp. 17–22.

Kwasnick, B., 1999. The Role of Classification in Knowledge Representation and Discovery. Library Trends 48 (1), 22–47.

Lee, B., Smith, G., Robertson, G., Czerwinski, M., Tan, D., 2009. FacetLens: Exposing Trends and Relationships to Support Sensemaking within Faceted Datasets. In: Olsen Jr., D., Arthur, R., Hinckley, K., Morris, M., Hudson, S., Greenberg, S. (Eds.), Proceedings of the Twenty-seventh ACM International Conference on Human Factors in Computing Systems (CHI). ACM Press, New York, NY, USA, pp. 1293–1302.

Manber, U., Smith, M., Gopal, B., 1997. WebGlimpse: Combining Browsing and Searching. In: Johnson, M. (Ed.), Proceedings of the USENIX Annual Technical Conference.

Marsden, G., Cairns, D., 2003. Improving the Usability of the Hierarchical File System. In: Eloff, J., Engelbrecht, A., Kotzé, P., Eloff, M. (Eds.), Proceedings of the Annual Research Conference of the South African Institute of Computer Scientists and Information Technologists on Enablement through Technology (SAICSIT). South African Institute for Computer Scientists and Information Technologists, Republic of South Africa, pp. 122–129.

McGuffin, M., m.c. schraefel, 2004. A Comparison of Hyperstructures: Zzstructures, mSpaces, and Polyarchies. In: Whitehead, J. (Ed.), Proceedings of the Fifteenth ACM Conference on Hypertext and Hypermedia (HT). ACM Press, New York, NY, pp. 153–162.

Medynskiy, Y., Dontcheva, M., Drucker, S., 2009. Exploring Websites through Contextual Facets. In: Olsen, Jr., D., Arthur, R., Hinckley, K., Morris, M., Hudson, S., Greenberg, S. (Eds.), Proceedings of the Twenty-seventh ACM International Conference on Human Factors in Computing Systems (CHI). ACM Press, New York, NY, pp. 2013–2022.

Merholz, P., Sep. 23, 2001. Innovation in Classification. URL: `http://www.peterme.com/archives/00000063.html`.

Meuss, H., Schulz, K., 2001. Complete Answer Aggregates for Treelike Databases: A Novel Approach to Combine Querying and Navigation. ACM Transactions on Information Systems 19 (2), 161–215.

Morville, P., Rosenfeld, L., 2006. Information Architecture for the World Wide Web, Third Edition. O'Reilly and Associates, Inc., Sebastopol, CA.

Narayan, M., Williams, C., Perugini, S., Ramakrishnan, N., 2004. Staging Transformations for Multimodal Web Interaction Management. In: Najork, M., Wills, C. (Eds.), Proceedings of the Thirteenth International ACM World Wide Web Conference (WWW). ACM Press, New York, NY, pp. 212–223.

Nevill-Manning, C., Witten, I., Paynter, G., 1999. Lexically-Generated Subject Hierarchies for Browsing Large Collections. International Journal on Digital Libraries 1 (2–3), 111–123.

Nielsen, J., 1990. Hypertext and Hypermedia. Academic Press, San Diego, CA.

Olston, C., Chi, E., 2003. ScentTrails: Integrating Browsing and Searching on the Web. ACM Transactions on Computer-Human Interaction 10 (3), 177–197.

Perkowitz, M., Etzioni, O., 2000. Adaptive Web Sites. Communications of the ACM 43 (8), 152–158.

Perugini, S., May 2004. Program Transformations for Information Personalization. Ph.D. thesis, Department of Computer Science, Virginia Tech.

Perugini, S., 2006. Real-time Query Expansion and Procedural Interfaces for Information Hierarchies. In: Broder, A., Maarek, Y. (Eds.), Proceedings of the International ACM SIGIR Workshop on Faceted Search.

Perugini, S., 2008. Symbolic Links in the Open Directory Project. Information Processing and Management 44 (2), 910–930.

Perugini, S., 2009. Personalization by Website Transformation: Theory and Practice. Revision under review for publication in *Information Processing and Management*.

Perugini, S., Ramakrishnan, N., 2003. Personalizing Web Sites with Mixed-Initiative Interaction. IEEE IT Professional 5 (2), 9–15.

Perugini, S., Ramakrishnan, N., 2007. Mining Web Functional Dependencies for Flexible Information Access. Journal of the American Society for Information Science and Technology (JASIST) 58 (12), 1805–1819.

Prieto-Díaz, R., 1991. Implementing Faceted Classification for Software Reuse. Communications of the ACM 34 (5), 88–97.

Ranganathan, S., 1951. Philosophy of Library Classification. E. Munksgaard, Copenhagen.

Rao, U., Turoff, M., 1990. Hypertext Functionality: A Theoretical Framework. International Journal of Human-Computer Interaction 2 (4), 333–358.

Robertson, G., Cameron, K., Czerwinski, M., Robbins, D., 2002. Animated Visualization of Multiple Intersecting Hierarchies. Information Visualization 1, 50–65.

Sacco, G., 2000. Dynamic Taxonomies: A Model for Large Information Bases. IEEE Transactions on Knowledge and Data Engineering 12 (3), 468–479.

Sacco, G., Tzitzikas, Y. (Eds.), 2009. Dynamic Taxonomies and Faceted Search: Theory, Practice, and Experience. Springer, Berlin.

Sinha, V., Karger, D., 2005. Magnet: Supporting Navigation in Semistructured Data Environments. In: Özcan, F. (Ed.), Proceedings of the ACM SIGMOD International Conference on Management of Data (SIGMOD). ACM Press, New York, NY, pp. 97–106.

Spiteri, L., 1998. A Simplified Model for Facet Analysis: Ranganathan 101. Canadian Journal of Information and Library Science 23 (1/2), 1–30.

Stoica, E., Hearst, M., Richardson, M., 2007. Automating Creation of Hierarchical Faceted Metadata Structures. In: Sidner, C., Schultz, T., Stone, M., Zhai, C. (Eds.), Proceedings of the Annual Conference of the North American Chapter of the Association for Computational Linguistics Human Language Technology Conference (NAACL-HLT). The Association for Computational Linguistics, Prague, Czech Republic, pp. 244–251.

Taylor, A., 1999. The Organization of Information. Libraries Unlimited, Inc., Englewood, CO.

Tvarožek, M., Bieliková, M., 2008. Collaborative Multi-Paradigm Exploratory Search. In: Huai, J., Chen, R., Hon, H.-W., Liu, Y., Ma, W.-Y., Tomkins, A., Zhang, X. (Eds.), Proceedings of the Hypertext Workshop on

Collaboration and Collective Intelligence (WebScience). ACM Press, New York, NY, pp. 29–33.

Tvarozek, M., Bielikova, M., 2008. Personalized View-based Search and Visualization as a Means for Deep/Semantic Web Data Access. In: Huai, J., Chen, R., Hon, H.-W., Liu, Y., Ma, W.-Y., Tomkins, A., Zhang, X. (Eds.), Proceedings of the Seventeenth ACM International World Wide Web Conference (WWW). ACM Press, New York, NY, pp. 1023–1024.

Tzitzikas, Y., Analyti, A., Spyratos, N., Constantopoulos, P., 2007. An Algebra for Specifying Valid Compound Terms in Faceted Taxonomies. Data and Knowledge Engineering 62 (1), 1–40.

Weiland, M., Dachselt, R., 2008. Facet Folders: Flexible Filter Hierarchies with Faceted Metadata. In: Czerwinski, M., Lund, A., Tan, D. (Eds.), Extended Abstracts on Human Factors in Computing Systems (CHI). ACM Press, New York, NY, pp. 3735–3740.

White, R., Marchionini, G., 2006. A Study of Real-time Query Expansion Effectiveness. In: Efthimiadis, E., Dumais, S., Hawking, D., Järvelin, K. (Eds.), Proceedings of the Twenty-ninth Annual International ACM Conference on Research and Development in Information Retrieval (SIGIR). ACM Press, New York, NY, pp. 715–716.

Wilson, M., André, P., m.c. schraefel, 2008. Backward Highlighting: Enhancing Faceted Search. In: Cousins, S., Beaudouin-Lafon, M. (Eds.), Proceedings of the Twenty-first Annual ACM International Symposium on User Interface Software and Technology (UIST). ACM Press, New York, NY, pp. 235–238.

Wilson, M., m.c. schraefel, 2008. A Longitudinal Study of Exploratory and Keyword Search. In: Larsen, R., Paepcke, A., Borbinha, J., Naaman, M. (Eds.), Proceedings of the Eighth International ACM/IEEE-CS Joint Conference on Digital Libraries (JCDL). ACM Press, New York, NY, pp. 52–56.

Wynar, B., 2000. In: Taylor, A. (Ed.), Waynar's Introduction to Cataloging and Classification, Ninth Edition. Libraries Unlimited, Inc., Englewood, CO.

Yee, K.-P., Swearingen, K., Li, K., Hearst, M., 2003. Faceted Metadata for Image Search and Browsing. In: Cockton, G., Korhonen, P. (Eds.), Proceedings of the ACM International Conference on Human Factors in Computing Systems (CHI). ACM Press, New York, NY, pp. 401–408.

Zhang, J., Marchionini, G., 2004. Coupling Browse and Search in Highly Interactive User Interfaces: a Study of the Relation Browser++. In: Chen, H., Wactlar, H., Chen, C., Lim, E.-P., Christel, M. (Eds.), Proceedings of the Fourth International ACM/IEEE-CS Joint Conference on Digital Libraries (JCDL). ACM Press, New York, NY, p. 384.

Tables

Table 1

	within	**between**
poorly connected	multiclassification links bridging top-level categories of top-level categories	
well connected	multiclassification links bridging topics sharing at least the first two levels of topic specificity	multiclassification links bridging top-level categories

Table 2

	faceted classification	**faceted search**	**symbolic links**
flexibility	$\surd$	$\surd$	$\surd$
disorientation	$\times$	some	$\surd$
context switch	$\times$	$\surd$	$\times$
enumeration	$\surd$	$\times$	some
anticipation	$\times$	$\times$	$\surd$
goal articulation	$\times$	$\surd$	$\times$

Figures

Figure 1

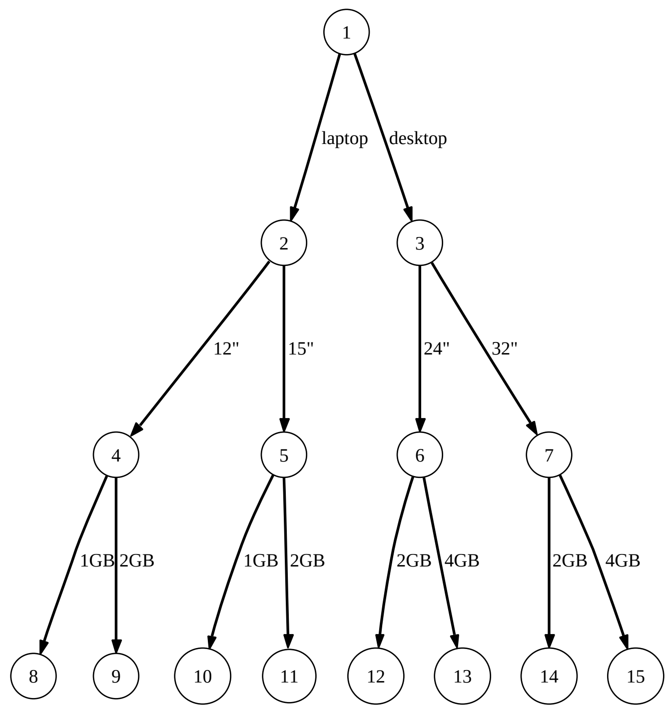

Figure 2

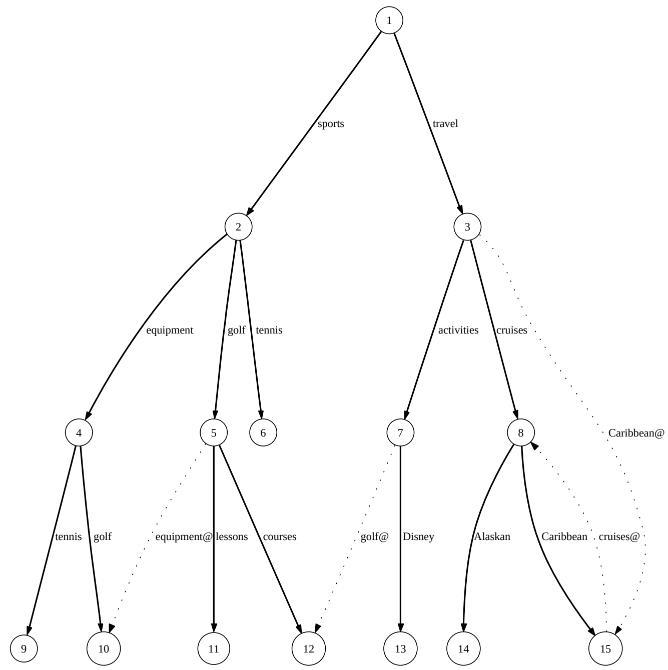
1
sports
travel
2
3
equipment
golf
tennis
activities
cruises
4
5
6
7
8
Caribbean@
tennis
golf
equipment@
lessons
courses
golf@
Disney
Alaskan
Caribbean
cruises@
9
10
11
12
13
14
15

Figure 3

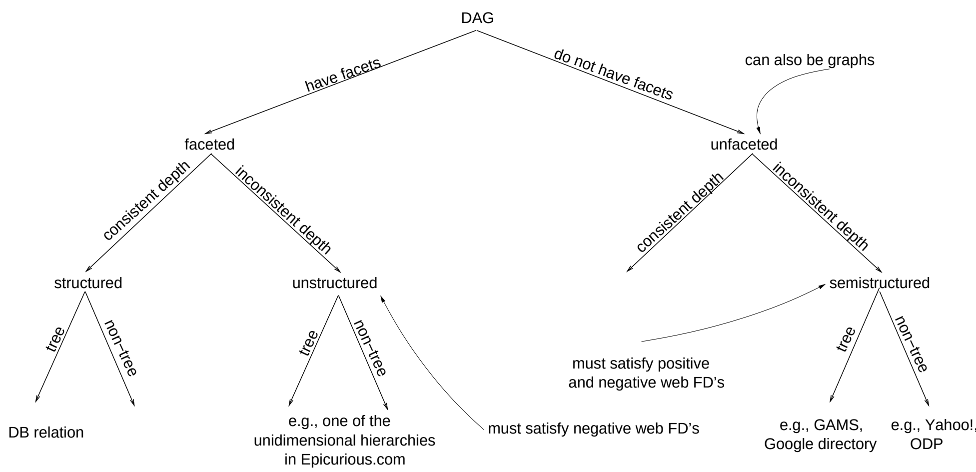
DAG
have facets
do not have facets
can also be graphs
faceted
unfaceted
consistent depth
inconsistent depth
consistent depth
inconsistent depth
structured
unstructured
semistructured
tree
non−tree
tree
non−tree
must satisfy positive
and negative web FD's
tree
non−tree
DB relation
e.g., one of the
unidimensional hierarchies
in Epicurious.com
must satisfy negative web FD's
e.g., GAMS,
Google directory
e.g., Yahoo!,
ODP

Figure 4

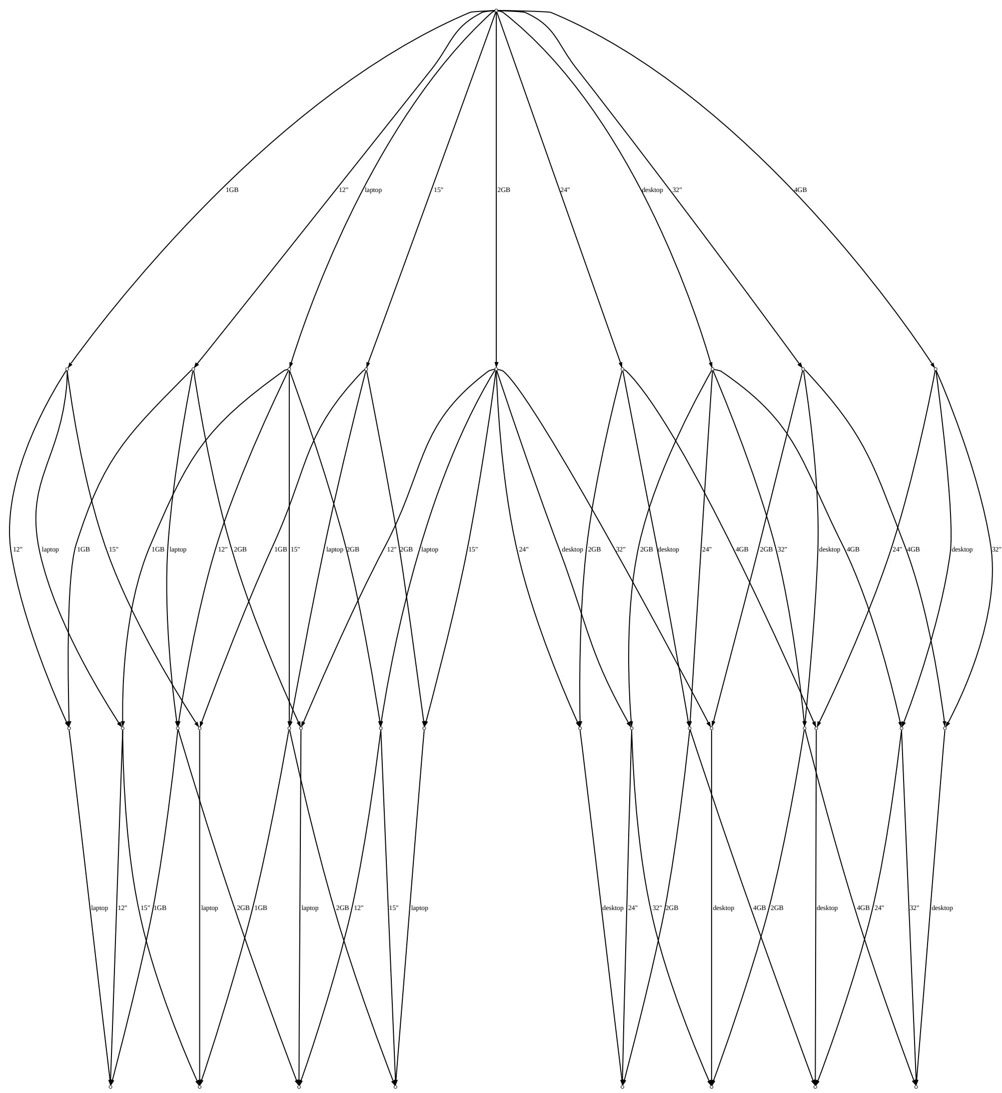

Figure 5

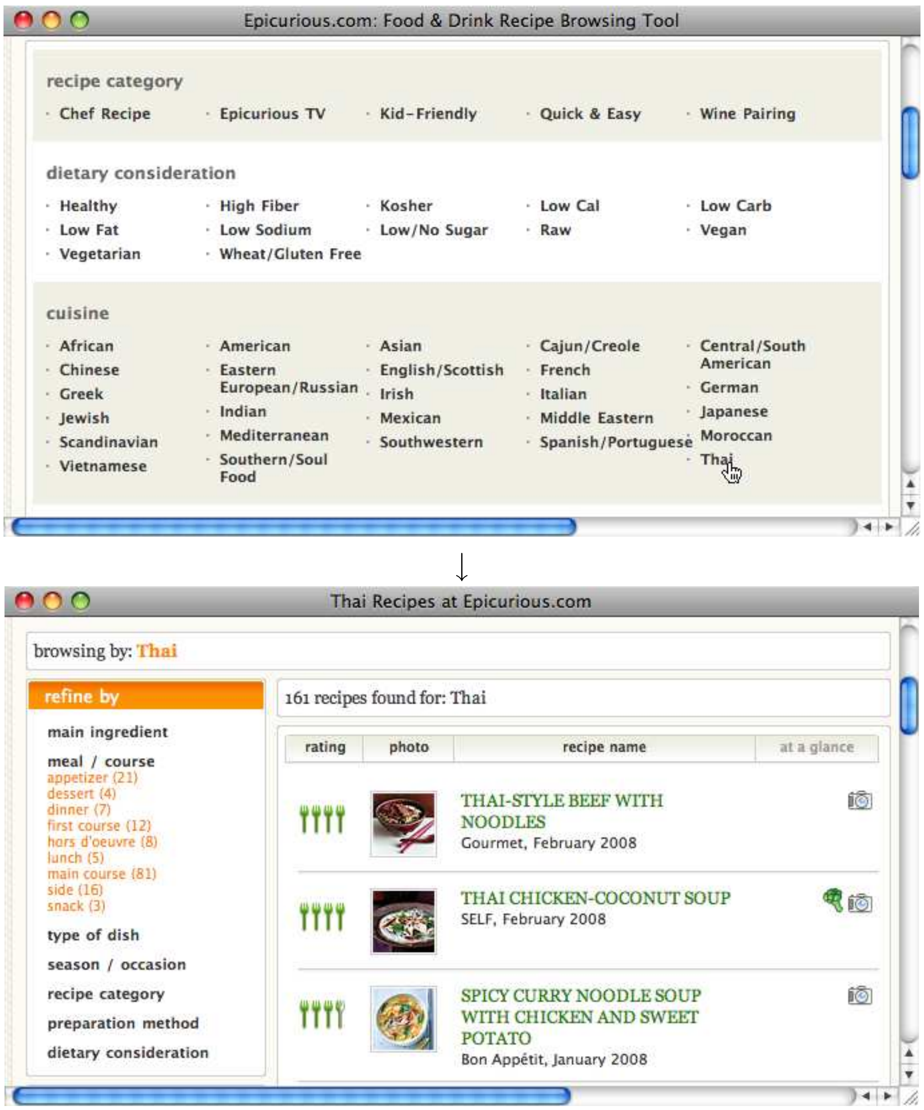
Epicurious.com: Food & Drink Recipe Browsing Tool
recipe category
Chef Recipe
Epicurious TV
Kid-Friendly
Quick & Easy
Wine Pairing
dietary consideration
Healthy
High Fiber
Kosher
Low Cal
Low Carb
Low Fat
Low Sodium
Low/No Sugar
Raw
Vegan
Vegetarian
Wheat/Gluten Free
cuisine
African
American
Asian
Cajun/Creole
Central/South American
Chinese
Eastern European/Russian
English/Scottish
French
German
Greek
Irish
Italian
Japanese
Jewish
Indian
Mexican
Middle Eastern
Moroccan
Scandinavian
Mediterranean
Southwestern
Spanish/Portuguese
Thai
Vietnamese
Southern/Soul Food
Thai Recipes at Epicurious.com
browsing by: Thai
refine by
main ingredient
meal / course
appetizer (21)
dessert (4)
dinner (7)
first course (12)
hors d'oeuvre (8)
lunch (5)
main course (81)
side (16)
snack (3)
type of dish
season / occasion
recipe category
preparation method
dietary consideration
161 recipes found for: Thai
rating
photo
recipe name
at a glance
THAI-STYLE BEEF WITH NOODLES
Gourmet, February 2008
THAI CHICKEN-COCONUT SOUP
SELF, February 2008
SPICY CURRY NOODLE SOUP WITH CHICKEN AND SWEET POTATO
Bon Appétit, January 2008

Figure 6

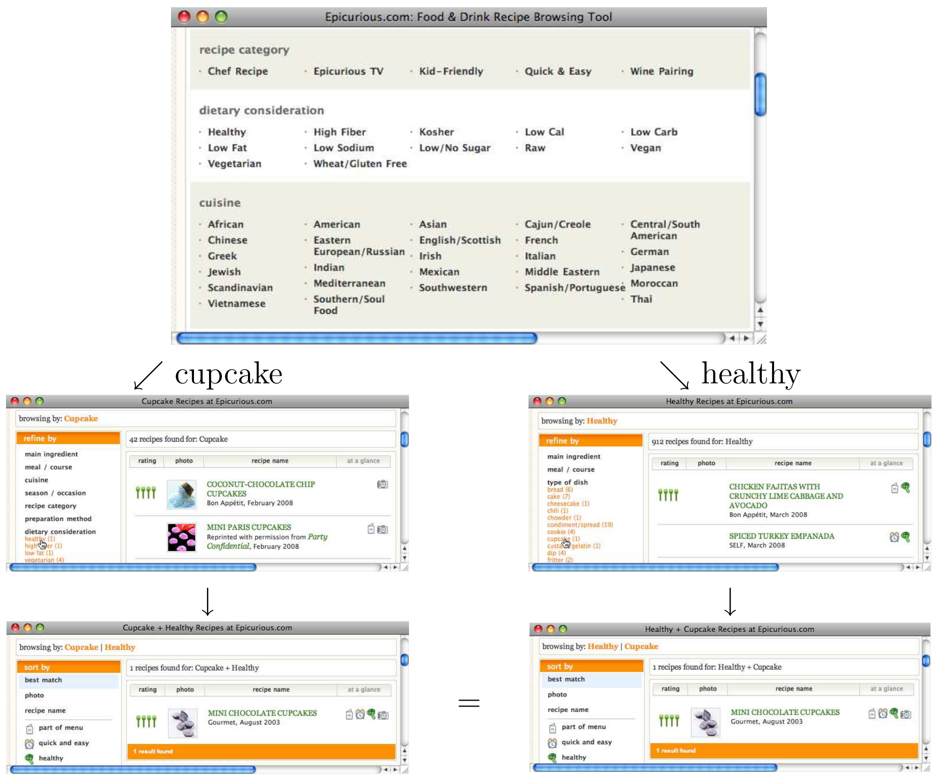
Epicurious.com: Food & Drink Recipe Browsing Tool
recipe category
Chef Recipe
Epicurious TV
Kid-Friendly
Quick & Easy
Wine Pairing
dietary consideration
Healthy
High Fiber
Kosher
Low Cal
Low Carb
Low Fat
Low Sodium
Low/No Sugar
Raw
Vegan
Vegetarian
Wheat/Gluten Free
cuisine
African
American
Asian
Cajun/Creole
Central/South American
Chinese
Eastern European/Russian
English/Scottish
French
German
Greek
Irish
Italian
Japanese
Indian
Jewish
Mexican
Middle Eastern
Moroccan
Mediterranean
Scandinavian
Southwestern
Spanish/Portuguese
Thai
Southern/Soul Food
Vietnamese
cupcake
healthy
Cupcake Recipes at Epicurious.com
browsing by: Cupcake
42 recipes found for: Cupcake
COCONUT-CHOCOLATE CHIP CUPCAKES
MINI PARIS CUPCAKES
Healthy Recipes at Epicurious.com
browsing by: Healthy
912 recipes found for: Healthy
CHICKEN FAJITAS WITH CRUNCHY LIME CABBAGE AND AVOCADO
SPICED TURKEY EMPANADA
Cupcake + Healthy Recipes at Epicurious.com
browsing by: Cupcake | Healthy
1 recipes found for: Cupcake + Healthy
MINI CHOCOLATE CUPCAKES
Gourmet, August 2003
1 result found
=
Healthy + Cupcake Recipes at Epicurious.com
browsing by: Healthy | Cupcake
1 recipes found for: Healthy + Cupcake
MINI CHOCOLATE CUPCAKES
Gourmet, August 2003
1 result found

Figure 7

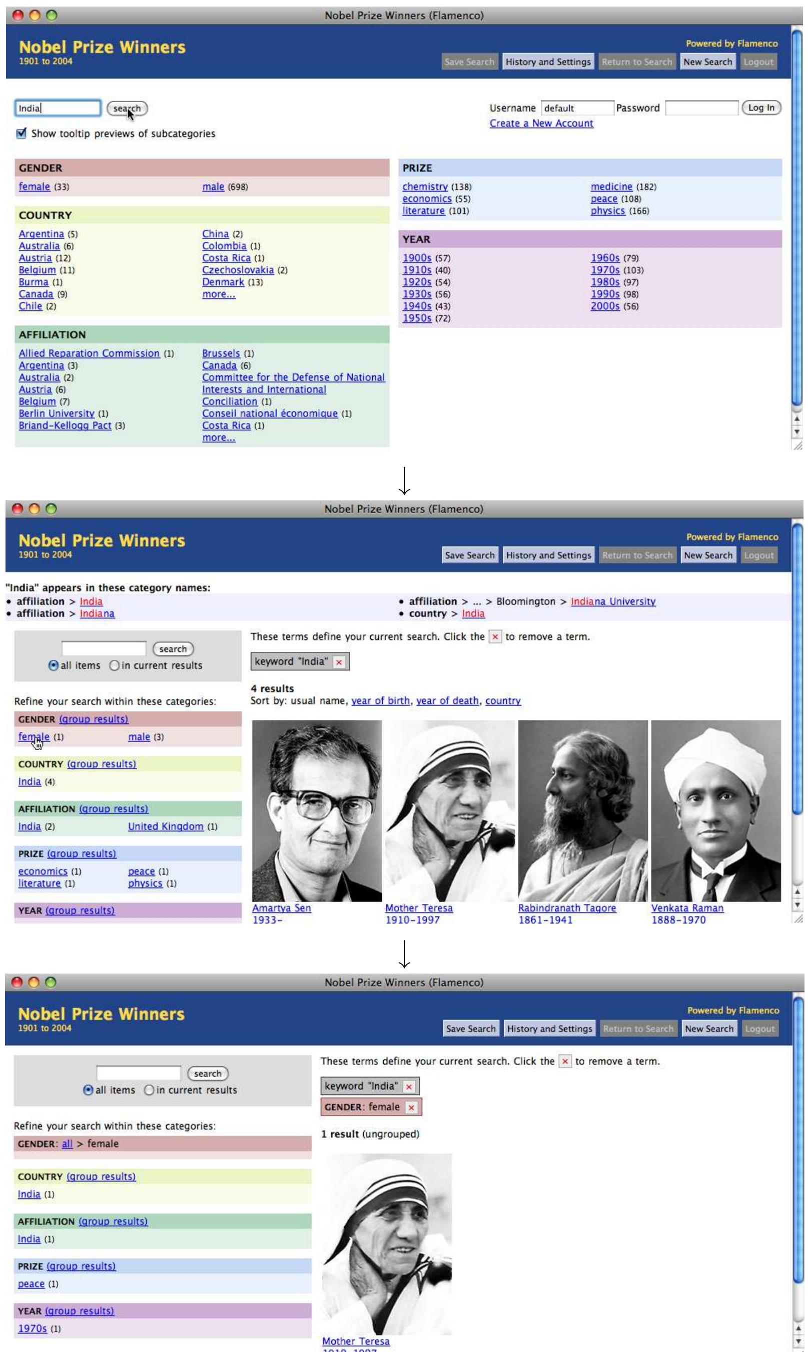
Nobel Prize Winners (Flamenco)
Nobel Prize Winners
1901 to 2004
Powered by Flamenco
Save Search
History and Settings
Return to Search
New Search
Logout
India
search
Username default Password
Log In
Create a New Account
Show tooltip previews of subcategories
GENDER
female (33) male (698)
COUNTRY
Argentina (5) Australia (6) Austria (12) Belgium (11) Burma (1) Canada (9) Chile (2)
China (2) Colombia (1) Costa Rica (1) Czechoslovakia (2) Denmark (13) more...
AFFILIATION
Allied Reparation Commission (1) Argentina (3) Australia (2) Austria (6) Belgium (7) Berlin University (1) Briand-Kellogg Pact (3)
Brussels (1) Canada (6) Committee for the Defense of National Interests and International Conciliation (1) Conseil national économique (1) Costa Rica (1) more...
PRIZE
chemistry (138) economics (55) literature (101)
medicine (182) peace (108) physics (166)
YEAR
1900s (57) 1910s (40) 1920s (54) 1930s (56) 1940s (43) 1950s (72)
1960s (79) 1970s (103) 1980s (97) 1990s (98) 2000s (56)
"India" appears in these category names:
affiliation > India
affiliation > Indiana
affiliation > ... > Bloomington > Indiana University
country > India
search
all items in current results
These terms define your current search. Click the × to remove a term.
keyword "India"
4 results
Sort by: usual name, year of birth, year of death, country
Refine your search within these categories:
GENDER (group results)
female (1) male (3)
COUNTRY (group results)
India (4)
AFFILIATION (group results)
India (2) United Kingdom (1)
PRIZE (group results)
economics (1) peace (1)
literature (1) physics (1)
YEAR (group results)
Amartya Sen
1933–
Mother Teresa
1910–1997
Rabindranath Tagore
1861–1941
Venkata Raman
1888–1970
GENDER: female
1 result (ungrouped)
GENDER: all > female
COUNTRY (group results)
India (1)
AFFILIATION (group results)
India (1)
PRIZE (group results)
peace (1)
YEAR (group results)
1970s (1)
Mother Teresa
1910–1997

Figure 8

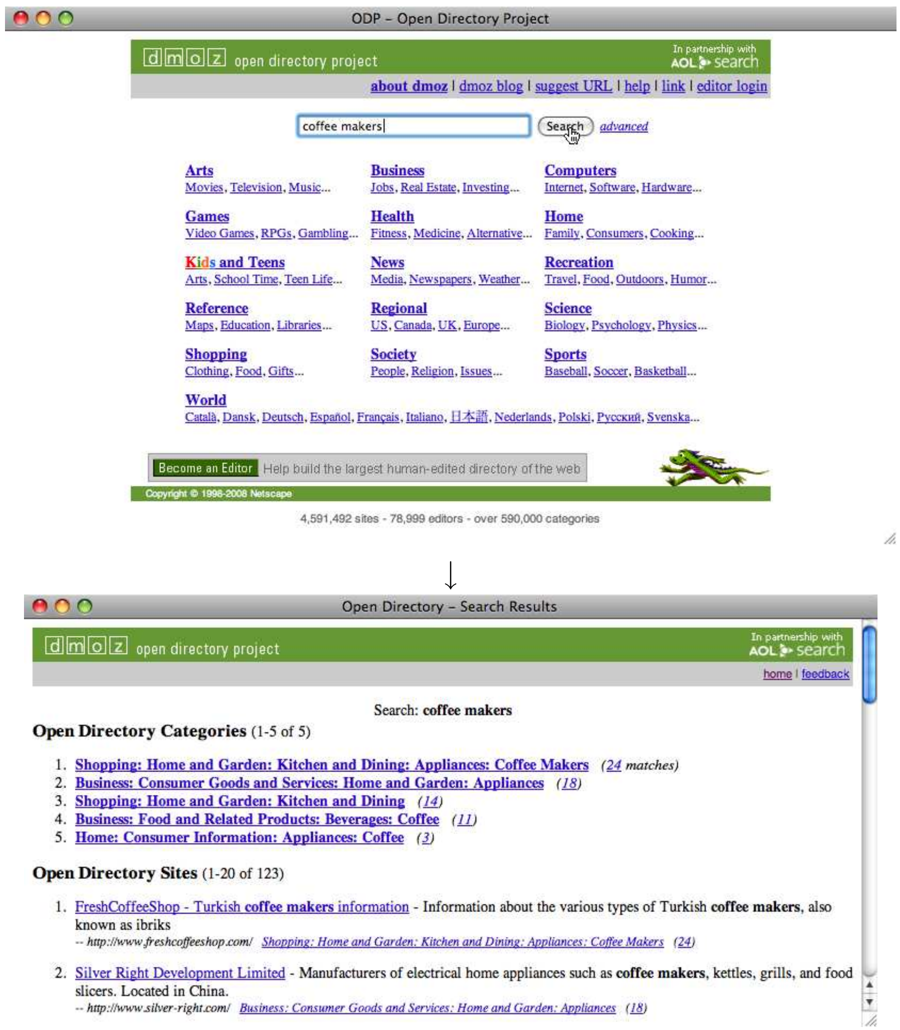
ODP – Open Directory Project
dmoz open directory project
In partnership with AOL search
about dmoz | dmoz blog | suggest URL | help | link | editor login
coffee makers
Search
advanced
Arts
Movies, Television, Music...
Business
Jobs, Real Estate, Investing...
Computers
Internet, Software, Hardware...
Games
Video Games, RPGs, Gambling...
Health
Fitness, Medicine, Alternative...
Home
Family, Consumers, Cooking...
Kids and Teens
Arts, School Time, Teen Life...
News
Media, Newspapers, Weather...
Recreation
Travel, Food, Outdoors, Humor...
Reference
Maps, Education, Libraries...
Regional
US, Canada, UK, Europe...
Science
Biology, Psychology, Physics...
Shopping
Clothing, Food, Gifts...
Society
People, Religion, Issues...
Sports
Baseball, Soccer, Basketball...
World
Català, Dansk, Deutsch, Español, Français, Italiano, 日本語, Nederlands, Polski, Русский, Svenska...
Become an Editor
Help build the largest human-edited directory of the web
Copyright © 1998-2008 Netscape
4,591,492 sites - 78,999 editors - over 590,000 categories
↓
Open Directory – Search Results
dmoz open directory project
In partnership with AOL search
home | feedback
Search: coffee makers
Open Directory Categories (1-5 of 5)
1. Shopping: Home and Garden: Kitchen and Dining: Appliances: Coffee Makers (24 matches)
2. Business: Consumer Goods and Services: Home and Garden: Appliances (18)
3. Shopping: Home and Garden: Kitchen and Dining (14)
4. Business: Food and Related Products: Beverages: Coffee (11)
5. Home: Consumer Information: Appliances: Coffee (3)
Open Directory Sites (1-20 of 123)
1. FreshCoffeeShop - Turkish coffee makers information - Information about the various types of Turkish coffee makers, also known as ibriks
-- http://www.freshcoffeeshop.com/ Shopping: Home and Garden: Kitchen and Dining: Appliances: Coffee Makers (24)
2. Silver Right Development Limited - Manufacturers of electrical home appliances such as coffee makers, kettles, grills, and food slicers. Located in China.
-- http://www.silver-right.com/ Business: Consumer Goods and Services: Home and Garden: Appliances (18)

Figure 9


ODP – Open Directory Project
dmoz open directory project
In partnership with AOL search
about dmoz | dmoz blog | suggest URL | help | link | editor login
coffee makers
Search
advanced
Arts
Movies, Television, Music...
Business
Jobs, Real Estate, Investing...
Computers
Internet, Software, Hardware...
Games
Video Games, RPGs, Gambling...
Health
Fitness, Medicine, Alternative...
Home
Family, Consumers, Cooking...
Kids and Teens
Arts, School Time, Teen Life...
News
Media, Newspapers, Weather...
Recreation
Travel, Food, Outdoors, Humor...
Reference
Maps, Education, Libraries...
Regional
US, Canada, UK, Europe...
Science
Biology, Psychology, Physics...
Shopping
Clothing, Food, Gifts...
Society
People, Religion, Issues...
Sports
Baseball, Soccer, Basketball...
World
Català, Dansk, Deutsch, Español, Français, Italiano, 日本語, Nederlands, Polski, Русский, Svenska...
Become an Editor
Help build the largest human-edited directory of the web
Copyright © 1998-2008 Netscape
4,591,492 sites - 78,999 editors - over 590,000 categories
↓
ODP – Open Directory Project
dmoz open directory project
In partnership with AOL search
about dmoz | dmoz blog | suggest URL | help | link | editor login
Search
advanced
Business
Home
Consumers
Shopping

Figure 10

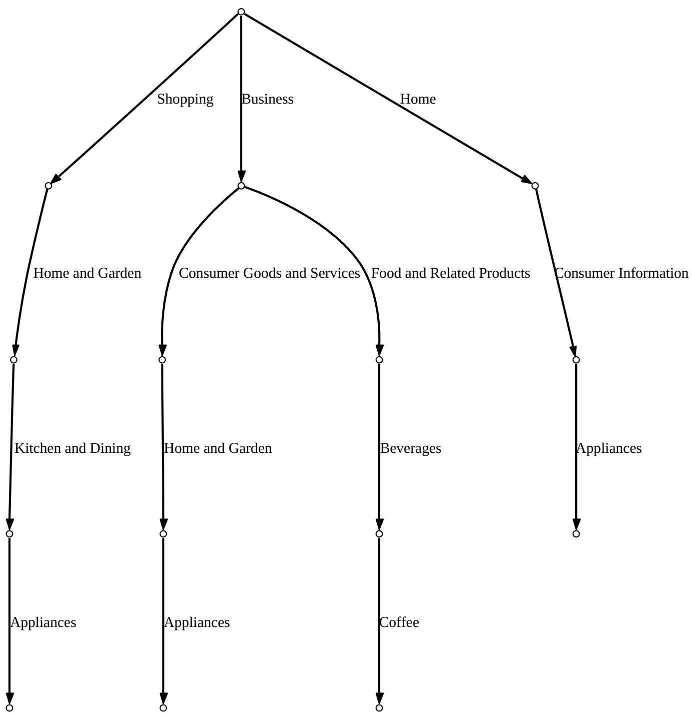
Shopping
Business
Home
Home and Garden
Consumer Goods and Services
Food and Related Products
Consumer Information
Kitchen and Dining
Home and Garden
Beverages
Appliances
Appliances
Appliances
Coffee

Figure 11

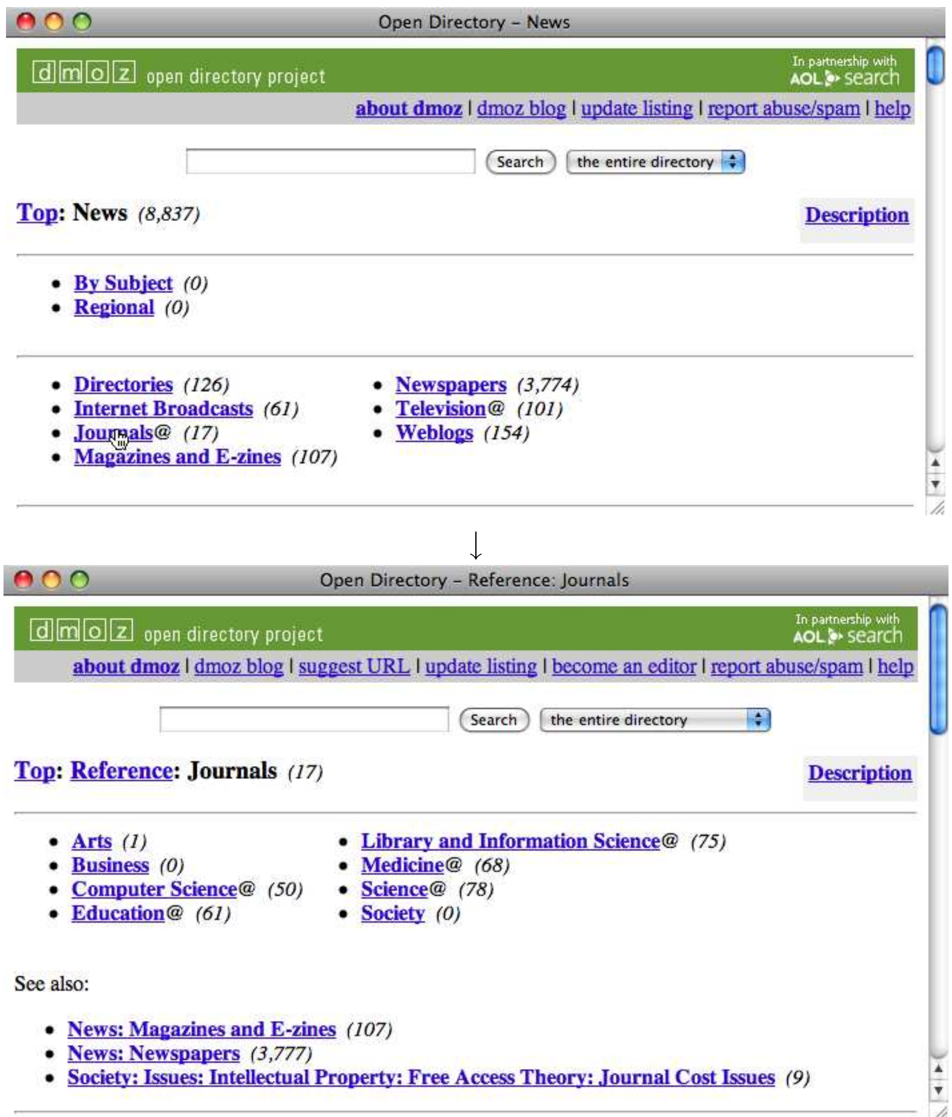
Open Directory – News
dmoz open directory project
In partnership with AOL search
about dmoz | dmoz blog | update listing | report abuse/spam | help
Search
the entire directory
Top: News (8,837)
Description
By Subject (0)
Regional (0)
Directories (126)
Internet Broadcasts (61)
Journals@ (17)
Magazines and E-zines (107)
Newspapers (3,774)
Television@ (101)
Weblogs (154)
↓
Open Directory – Reference: Journals
dmoz open directory project
In partnership with AOL search
about dmoz | dmoz blog | suggest URL | update listing | become an editor | report abuse/spam | help
Search
the entire directory
Top: Reference: Journals (17)
Description
Arts (1)
Business (0)
Computer Science@ (50)
Education@ (61)
Library and Information Science@ (75)
Medicine@ (68)
Science@ (78)
Society (0)
See also:
News: Magazines and E-zines (107)
News: Newspapers (3,777)
Society: Issues: Intellectual Property: Free Access Theory: Journal Cost Issues (9)

Figure 12

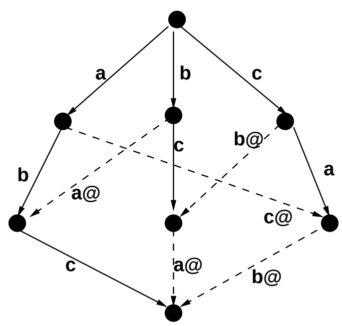

Figure 13

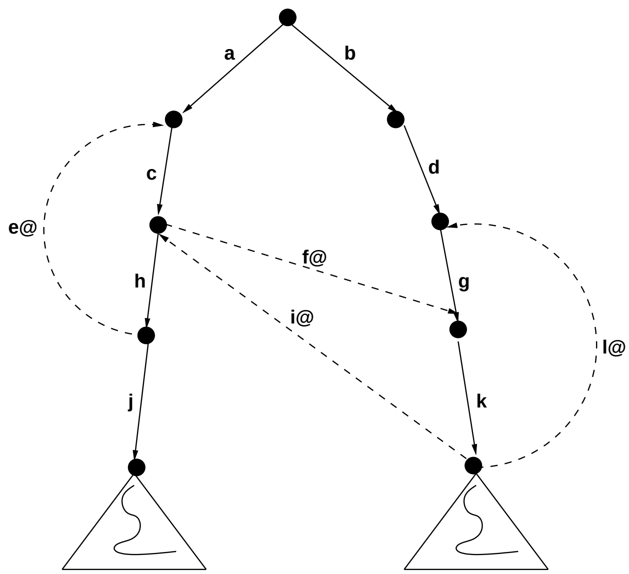

Figure 14

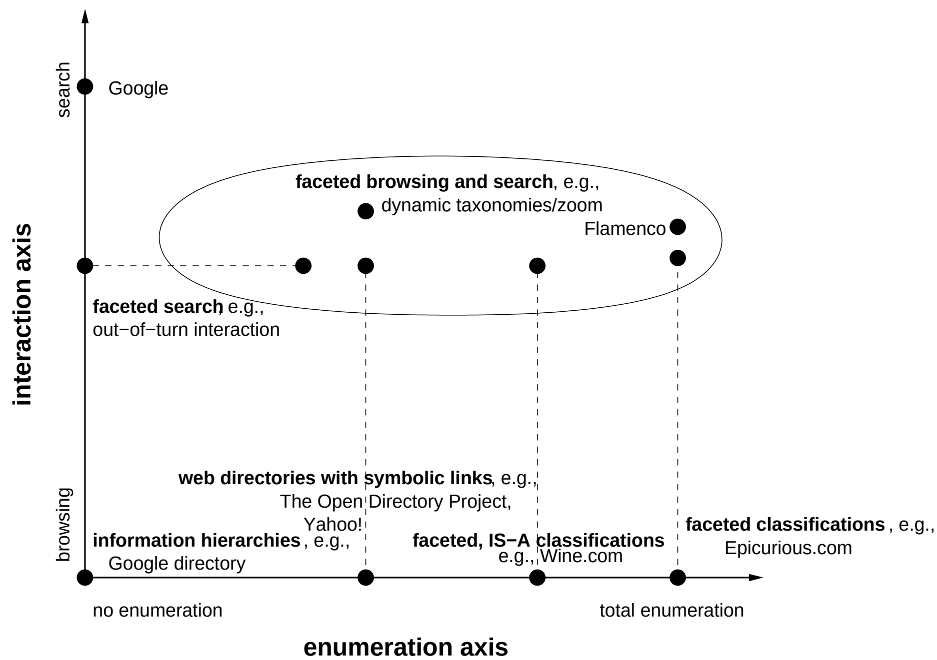
search
Google
faceted browsing and search, e.g.,
dynamic taxonomies/zoom
Flamenco
interaction axis
faceted search, e.g.,
out−of−turn interaction
web directories with symbolic links, e.g.,
The Open Directory Project,
Yahoo!
browsing
information hierarchies, e.g.,
Google directory
faceted, IS−A classifications
e.g., Wine.com
faceted classifications, e.g.,
Epicurious.com
no enumeration
total enumeration
enumeration axis

Figure 15

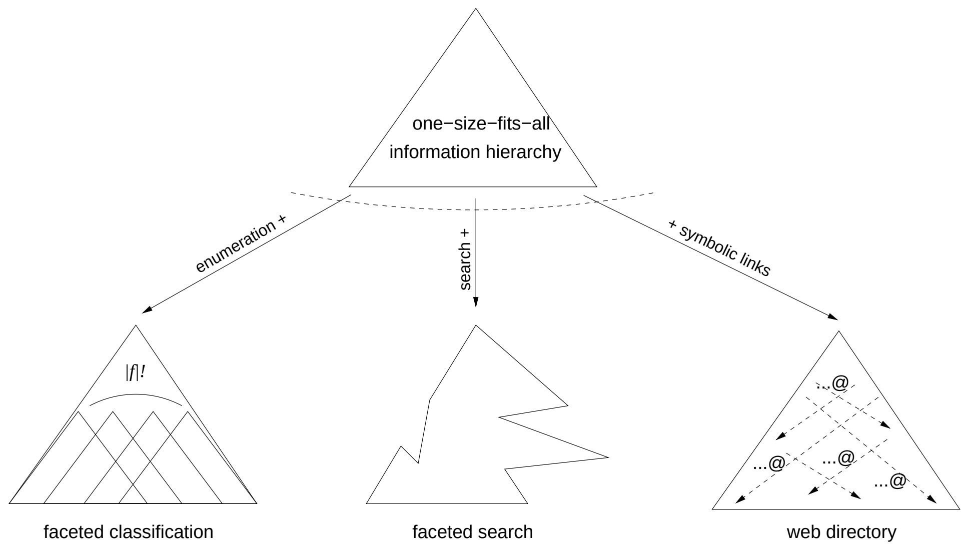
one−size−fits−all
information hierarchy
enumeration +
search +
+ symbolic links
|f|!
...@
...@
...@
...@
faceted classification
faceted search
web directory

Table captions

Caption for Table 1

Summary of results of (Perugini, 2008).

Caption for Table 2

A comparison of faceted classification, faceted search, and symbolic links across various attributes.

Figure captions

Caption for Figure 1

Sample structured web hierarchy simplified for purposes of presentation.

Caption for Figure 2

Example semistructured web directory sharing characteristics with *Yahoo!*. Symbolic links are indicated by dashed edges and hyperlink labels ending with @. All other typographical conventions follow those used in Fig. 1.

Caption for Figure 3

Classification of simple hierarchies described in this section.

Caption for Figure 4

A sample faceted classification with characteristics similar to those in Epicurious.com. Uses the same typographical conventions as the DAG in Fig. 1.

Caption for Figure 5

Faceted classification of recipes at Epicurious.com. The user has the option to choose a value for any facet remaining unfilled at any point in the interaction.

Caption for Figure 6

Path permutations in Epicurious.com. While both paths start at the same root (top) and end at the same recipe, the page on the bottom left is retrieved by following the $\prec$cupcake, healthy$\succ$ path and the page on the bottom right is accessed through the $\prec$healthy, cupcake$\succ$ path.

Caption for Figure 7

A sequence of interactions using the *Flamenco* faceted browsing and search interface.

Caption for Figure 8

A search for 'coffee makers' using the ODP search facility. (top) Root page of the ODP directory with a query in the search textfield. (bottom) Results of the search.

Caption for Figure 9

A search for 'coffee makers' in ODP using out-of-turn interaction. See Fig. 10 for the entire resulting hierarchy.

Caption for Figure 10

Schematic results, simplified for purposes of presentation, of a search for 'coffee makers' in ODP using out-of-turn interaction. Notice that the semblance of a hierarchy is preserved.

Caption for Figure 11

Illustration of a multiclassification symbolic link in the *Open Directory Project*. The user follows a symbolic link labeled 'Journals@' (top) which transports the user from the 'News' category (top) to the 'Reference: Journals' category (bottom).

Caption for Figure 12

Sample (factored) faceted classification, simplified for purposes of presentation, illustrating permuted paths induced by symbolic links. Symbolic links help realize all 6 (=3!) permutations of the one and only $\prec$a, b, c$\succ$ hard path (see leftmost path). Uses the same typographical conventions as Fig. 2.

Caption for Figure 13

Sample semistructured web directory contrasting *backlinks* and *backjumps*. Uses the same typographical conventions as Fig. 2.

Caption for Figure 14

Graphical overview of tradeoffs.

Caption for Figure 15

Approaches to supporting multiple access paths to each terminal object in information hierarchies. (left) information hierarchy + enumeration = faceted classification. (center) information hierarchy + search = faceted search. (right) information hierarchy + symbolic links = web directory.