G

Guided Product Selection and Comparison of E-Commerce Portals

Giovanni M. Sacco
Università di Torino, Italy

INTRODUCTION

The number of e-commerce sites is growing at an astounding rate. Low personnel overhead, just-in-time supply, and the widespread acceptance of online credit card payments make large e-stores a viable business model. Indeed, the economics of small or no inventories seems to be a compelling force toward mammoth stores; successful stores such as Amazon have shifted from a focused line of products to selling quite diverse and heterogeneous items.

E-commerce portals are one of the most active and important Internet application areas, yet selecting a product to buy is frequently a frustrating experience because of the size of inventory, and, most importantly, because customers do not know exactly the specific item they want, but are rather looking for the item that best fits their individual requirements. This situation requires system assistance in browsing and exploration as opposed to retrieval based on a precise specification, which is the paradigm of search tools (text retrieval and database queries) supplied by traditional technology and used in most portals.

BACKGROUND: THE USER ACCESS MODEL

Most users of e-commerce portals do not look for a specific product but want to find the "right" product in possibly a quite large set of alternative products. The right product really depends on how competing features rate according to user requirements (perceptions, interests, financial capabilities, etc.). Different users or even the same users at different times are likely to weigh each feature differently. While it is unlikely that users are able to associate a precise numeric weight to each feature, they can very easily rank features in decreasing order of importance. So, in addition to a primary interest focus (e.g., budget price), users will have a secondary, tertiary, etc. focus (e.g., budget cameras with the highest resolution vs. the lightest budget cameras available). A secondary focus depends on the user preferences but also on the features that items in the primary focus exhibit, and so on.

For expediency, we split the interaction into two stages in cascade: the thinning-game and the end game (Sacco, 2003, 2005). In the thinning game, the user is confronted with a large number of items and has to derive a relatively small set of candidate items to be further exhaustively inspected. In order to thin the number of alternatives the user has to:

1. find all the available features;
2. focus on the most relevant one for him (the primary focus) and discard all the items without that feature;
3. find all the features for the items retained; and
4. select the next focus among them and iterate the process until the number of candidates is sufficiently small.

The major critical point is the display of all the features correlated with the selected ones. What are the features (e.g., resolution, zoom, etc.) for cameras under $200? If the user is not able to find them out easily, the next focus cannot be set and the thinning game is already over. The user has to inspect all the cheap cameras and find their features by manual inspection. On the other hand, if related features are available, he or she can add the next feature in the order of perceived importance to the current focus and focus on it, thereby discarding other documents that do not have that feature and consequently further thinning the number of candidate items.

Other important points for the thinning game are the ability to operate on items at a set-at-a-time rather than at an instance-at-a-time level (the primary focus defines a set of items, a secondary focus intersects the primary focus set with the set defined by the secondary focus, etc.), and to have systematic summaries of sets (the current focus) in real time. Finally, the number of features for large stores can be quite large, so that a taxonomic organization of features is usually required. Item presentation tends to be a second-order concern in the thinning game.

The second stage, the end game, is entered when a suitably small set of candidate items has been located and the user must select the single item to purchase, by comparing features of candidate items. Candidate items are usually organized as a table with features on the rows and items on the columns. Feature comparison poses significant cognitive challenges to users because there are usually many features to consider and the number of candidate items is often larger than 10. Most practical situations may require hundreds of comparisons, but even the comparison of two items can be difficult--not only all the features must be compared, but all the different features must also be remembered. Different features will be stored in the user short-term memory (Miller, 1956), which holds 7±2 items. This means that comparing more than nine feature values becomes quite complex and usually leads to total user disorientation so that users will need additional tools such as pencil and paper.

The number of comparisons to be performed has to be minimized. Consequently, the user should be assisted in quickly finding *discriminants* among different items (i.e., features with different values that can guide the selection). At the very minimum, features whose values are constant over all the items, and are therefore useless as discriminants, should be quickly perceived as such, and discarded on demand. In addition, the user selects the final item user by informally *weighing* the desirableness of a combination of features of interest. In many practical cases, values of specific features can be ranked a priori from the less desirable to the most desirable value. For instance, being all the other features equal, a smaller price tag is always better than a higher one. These rankings can be used in such a way that the user quickly perceives the desirableness of feature values in a row, instead of comparing them exhaustively.

SOLUTIONS FOR PRODUCT SELECTION

Solutions for the Thinning Game

Shopping portals have used a number of different techniques to solve the thinning game. These include: (a) database queries, (b) text retrieval, (c) hypertext/hypermedia, and (d) taxonomies. For each technique, there are real or perceived system advantages. Most portals rely on relational technology for operation (inventory, ordering, billing, etc.) so that a form-based query system that operates on the underlying database requires a limited implementation effort. Text retrieval solutions require even less design and implementation. It is sufficient to index product sheets and they are immediately available. Hypertext/hypermedia can provide some sort of navigation in the inventory and are often used to implement static Yahoo-like taxonomies.

From the user perspective, none of these techniques satisfies the requirements of the thinning game. Database queries show lengthy result lists with no semantic structuring. They are good for precise retrieval but extremely poor for browsing. Setting the primary focus is easy; users interested in budget cameras just ask for cameras below $200 and retrieve a number of cameras that satisfy that condition. However, no summary of the features these cameras have is available, so that a secondary focus (e.g., small cameras or high resolution) cannot be set. Either users must read all the camera descriptions or they get involved in a lengthy trial-and-error interaction, issuing blind queries in order to find interesting features. Text retrieval queries are even worse. Noise and insufficient recalls are well known problems with text retrieval (Blair & Maron, 1985); queries on full-text material tend to retrieve too many documents or too few. For this reason, text retrieval is rarely used per se but rather in combination with database techniques as a way to access full-text descriptions of products.

Only the smallest e-stores can use hypertext/hypermedia techniques (Groenbaek & Trigg, 1994) for product selection. Although hypermedia is commonly used to browse information, exploration is performed one document at a time, which is quite time consuming, and there is no systematic picture of relationships among infobase components. Building and maintaining a complex hypermedia network can be extremely costly. The hierarchical topic structure of traditional taxonomies (such as Yahoo!) gives users an initial guidance and setting a primary focus is simple, but once they select a branch in the taxonomy, say price, the result can only be refined by descendant topics (i.e., a specific price range) and the discriminative power of features on other branches (i.e., resolution, weight, etc.) is lost. There is no way of setting a secondary focus and the only way to go on is really to manually inspect all the items in the primary focus. Note that this is not just a problem in designing the taxonomy, although one could replicate all the independent branches at each level (e.g., under *Zoom> optical zoom > 3X*, we could find *Price Range*, *Max. Resolution*, etc.), this strategy produces an exponential growth in the taxonomy.

A significant number of shopping portals based on dynamic taxonomies (see the article "Dynamic taxonomies: Intelligent user-centric access to complex portal information" in this Encyclopedia) have appeared in the past two years. These include, among others, Yahoo!, Kelkoo, Bizrate, and Amazon. Dynamic taxonomies are based on a multidimensional taxonomy in which items are classified under several concepts, and offer a single, integrated visual environment for retrieval and guided exploration. In the simplest case, the user can *zoom* on a concept C of interest; only the documents classified under C are retained, and all the concepts not related to the current focus are pruned from the taxonomy, which therefore shows all and only those concepts that can be used to set an additional focus. The term dynamic is used to indicate that dynamic taxonomies can conceptually summarize any subset of the universe, whereas traditional, static taxonomies are able to summarize only the entire universe.

The selection process is exemplified by a digital camera shop in Figures 1 to 3. The example uses Knowledge Processors' Universal Knowledge Processor (Knowledge

Related Content