Children are increasingly using the Web. Cognitive theory tells us that directory structures are especially suited for information retrieval by children; however, empirical results show that they prefer keyword searching. One of the reasons for these findings could be that the directory structures and terminology are created by grown-ups. Using a card-sorting method and an enveloping system, we simulated the structure of a directory. Our goal was to try to understand what browsable, hierarchical subject categories children create when suggested terms are supplied and they are free to add or delete terms. Twelve groups of four children each (fourth and fifth graders) participated in our exploratory study. The initial terminology presented to the children was based on names of categories used in popular directories, in the sections on Arts, Television, Music, Cinema, and Celebrities. The children were allowed to introduce additional cards and change the terms appearing on the 61 cards. Findings show that the different groups reached reasonable consensus; the majority of the category names used by existing directories were acceptable by them and only a small minority of the terms caused confusion. Our recommendation is to include children in the design process of directories, not only in designing the interface but also in designing the content structure as well.

Introduction

As an inseparable part of the information society, children are showing a growing interest in the advantages that the Internet has to offer over more traditional information sources such as books and encyclopedias. Children enjoy using the Net for information retrieval, communication, fun, and learning, mostly due to its accessibility and graphical richness.

Even though children are usually assumed to be highly technologically oriented, and children often perceive even themselves as such, research has shown that they experience a wide range of difficulties while searching for information on the Web. Existing Web search tools for children such as Yahooligans! (http://www.yahooligans.com), KidsClick (http://kidsclick.org), and Ask Jeeves for Kids (http://www.ajkids.com) do not contribute to children’s success in their information tasks. In fact, it has been found that these tools often cause a search failure due to their lack of special features essential for their young audience (Bilal, 1999, 2000, 2001, 2002a, Bilal & Watson, 1998; Large, Beheshti, & Rahman, 2002). One possible explanation for this lack of suitability to children’s needs can be the fact that search tools for children are mostly designed by adults, who do not actively include the target audience in the design process (Nesset & Large, 2004).

A few researchers have recently started to include children as equal partners in the design process of Web search tools in order to produce tools that suit children’s knowledge level and cognitive and conceptual levels. However, this approach is only in its primary stage and needs additional academic attention.

A Web directory is a browsable, hierarchical list of subjects that is used to index Web sites according to their content. Browsing, which is “an interactive process of skimming over information and selecting choices” (Borgman, Hirsh, Walter, & Gallagher, 1995), was found to be an easy and effective information searching method for children, because it requires less cognitive effort from them than recalling search terms from memory; therefore, children can concentrate better on finding the desired information.

This study concentrated on one of the more content-related aspects of the design process of Web search tools for children: the structure and terminology of the subject categories available in children’s Web directories. Using the “children as designers” approach, the purpose of this study was to investigate the process of categorization of information as perceived by elementary school children. We tried to gain a comprehensive understanding of the structure and terminology of subject categories that children would like to encounter in Web directories specially designed for them. The findings and the methodology of this study have implications...
on the content design of Web directories for children, which is a first step towards improving children’s online information retrieval success.

**Literature Review**

Over the last few years, children have become increasingly exposed to the Internet as an information source. In general, research has shown that children are interactive information seekers (Schacter, Chang, & Dorr, 1998) who do not tend to preplan their searches (Marchionini, 1989). They also do not tend to save their previous searches or favorite results and therefore often spend a long time trying to restore successful searches (Hirsh, 1999). Children were also found to be very persistent seekers (Bilal, 2000; Bilal & Watson, 1998) who are not likely to abandon their search task without at least a few tries to succeed in it (Borgman et al., 1995; Bowler, Large, & Rejskind, 2001; Large & Beheshti, 2000). However, they are not always quite aware of the fact that their queries might need to be improved, and they therefore sometimes spend a long time repeating unsuccessful searches over and over again (Large & Beheshti, 2000).

As for information-seeking techniques, evidently children are able to use browsing techniques effectively (Bilal, 2000; Borgman et al., 1995), and they are more successful in accomplishing their search tasks when browsing menus, categories, and hierarchies than when searching with keywords and Boolean operators (Bilal & Watson, 1998; Marchionini, 1989; Schacter, Chang & Dorr, 1998).

Children’s success with browsing information systems is explained by Borgman et al. (1995) based on cognitive theory. According to this theory, browsing information requires less cognitive effort from the child than recalling search terms from memory; therefore, the child can concentrate better on finding the desired information. Revelle et al. (2002) developed a visual search interface for retrieving information from a hierarchical information structure. This interface integrated searching with browsing and was tested on 5- to 10-year-old children.

While browsing reduces the child’s cognitive load, keyword searching is more challenging for a number of reasons. First, choosing the right keywords that express one’s information need is a very difficult task even for the adult seeker, and it is even more problematic for children (Large & Beheshti, 2000; Schacter, Chang, & Dorr, 1998). Children tend to use either too broad or too specific search terms, a fact that often prevents them from retrieving relevant information (Akin, 1998; Bilal & Watson, 1998). Children often submit natural language queries, a feature usually not supported by existing Web search tools (Bilal & Watson, 1998; Bilal, 2002a). The second cause of keyword searching failure is spelling mistakes, a very common problem among elementary school students (Bilal & Watson, 1998). And finally, applying the Boolean logic by using Boolean operators in the search query was found to be problematic for many elementary school children (Nahl & Harada, 1996). Nahl and Harada also found that children do not use recall to generate terms because their vocabularies are developing and incomplete for search language purposes.

Surprisingly, in spite of the fact that keyword searching seems to confront them with a wide range of difficulties and that browsing might be more suitable for their cognitive abilities, children seem to prefer applying keyword searching over browsing, especially as the first step of their Web searching process (Bilal & Watson, 1998; Belous & Baruchson-Arib, 2002; Large, Beheshiti, & Moukddad, 1999). Possible explanations for this kind of behavior can be associated with the problematic structure and terminology of the browsable categories and hierarchies available in existing Web search tools.

**Web Search Tools for Children**

Most available Web search tools are designed for the use of adult searchers. However, there are a few Web search tools specially designed for children. It could be expected that the Web search tools especially designed for children that address their Web search behavior will improve their success rate in online information searching tasks; however, research has shown that this is not always the case.

In her various studies of children’s use of Yahooligans, a browsable, searchable directory of Internet sites for kids aged 7–12 (Yahooligans, 2004), Bilal found that its system design does not contribute to children’s success with their information tasks. The lack of a misspelling checker, a thesaurus, a natural language interface, a rich school-related database, and user-friendly feedback affected the children’s low level of success (Bilal & Watson, 1998). Bilal also found that Yahooligans’ poor indexing and abstracting, which does not well reflect the content of the Web sites included in it, and the lack of context-sensitive online help, basic examples for searching and browsing, appropriate screen display, and individual display of categories and Web sites, decreased the children’s success rates and satisfaction (2000; 2001; 2002a). Other children-targeted Web search tools such as Ask Jeeves for kids, Super Snooper, KidsClick, and LycosZone were also not found to be quite as child-adapted as expected (Bilal, 1999; Large, Beheshiti, & Rahman, 2002).

One possible explanation for this lack of suitability to children’s needs may be the fact that search tools for children were mostly designed by adults, without actively including the target audience in the design process (Nesset & Large, 2004). Involving children as equal partners in the tool design process would result in the development of user-based search tools that conform to the cognitive and behavioral characteristics of children. This approach, as innovative as it might be, is already employed by software developers and HCI professionals in the course of the last few years. They feel that in order to design usable and effective technology products for children, children must be included in the design process itself and not only as users or testers of the finished product (Nesset & Large, 2004). Alison Druin was one of the pioneers who implemented this approach, starting
from the KidPad project in 1997 and later in her children’s digital library project (Druin, 2002; Druin et al., 1999; Druin et al., 2001).

**Children as Web Search Tool Designers**

One very early attempt of active involvement of children in the design process can be found in the SNAPdragon project (Kafai & Bates, 1997). Even though the main goal of this project was to “build children’s literacy skills,” the way in which this was achieved was quite unique. Elementary school children were asked to build an annotated Web site directory for other children. The researchers found that young children (grades 1–4) had trouble with evaluating Web sites and writing annotations. The children were very proud and enthusiastic when their directory went online and was accessible for others to see. This project, however, was not completely handled by the children, as the researchers admit; and, the adults were finally the ones having to decide how to categorize and index the sites and how to call the categories, so that eventually “the Dewey Decimal Classification main categories were adopted. In retrospect, the researchers wished that the children had had the opportunity to contend with the same issues” (Kafai & Bates, 1997). Indeed, adopting the Dewey Decimal Classification may not have been the most suitable selection for children’s use in this case, because children have been found to have a low level of understanding of Dewey’s categorization and terminology, especially when facing abstract categories (McMillan, 2000).

Belous and Baruchson-Arib (2002) took a step forward during the school year of 2000–2001 in an elementary school in Israel. In this research, 4th graders were also asked to build a Web site directory for the use of their friends; except this time, the children were not only assigned to collect suitable Web sites for their directory and annotate them but also asked to take responsibility for every other aspect of the project. Such aspects included identifying their friends’ information needs, retrieving the appropriate sites, organizing them, categorizing them, storing them, and eventually publishing the project online. Involvement of the adults in the project was limited to teaching the students how to use search engines effectively in order to retrieve Web sites and to translate the instructions (the children were not English speakers) for uploading the site to the free Web server (Tripod), which was used in the project. This study found that allowing children to deal with the whole process of building a Web search tool led to two important outcomes. First, children became very responsible information handlers who were constantly concerned with the quality of the information both retrieved and published. Second, the structure of the directory and the terminology used for the categories was more suitable to children’s cognitive state and information needs than existing directories created by and for adults.

A different approach to including children in Web search tools design process was taken by Andrew Large and his colleagues. Inspired by Druin’s experience, this team started by conducting a focus group user study in which elementary school students were encouraged to express their likes and dislikes regarding the interface design of four existing Web portals for children: Ask Jeeves for Kids, KidsClick, Lycos Zone and Yahooligans (Large, Beheshti, & Rahman, 2002). This focus group study revealed that the database of the Web search tool for children should cover both educational and entertainment related topics, have a catchy name and URL, and an eye-catching graphic design, as well as a very clear and direct results layout. They should also include both keyword searching and subject categories, preferably without a very deep hierarchical structure. Finally, the existence of personalization options was also suggested.

After clarifying children’s attitudes towards existing Web portals, Large and his colleagues established two intergenerational portal design teams, each comprised of elementary school students and adult researchers as equal design partners. The teams’ goal was to design a low-tech portal prototype for the use of elementary school students (Large, Beheshti, Nnesset, & Bowler, 2003, 2004). This unique design process allowed the researchers to gain very interesting insights into children’s portal design preferences as well as the effectiveness of intergenerational groups in children’s portal design. First, it was revealed that, when children were given the opportunity to design a portal of their own, they suggested quite a different design than usually accepted in adults’ portals—even though most of the children used adults’ Web portals in everyday life and even preferred them over available Children’s portals (such as Yahooligans or KidsClick). Children obviously have a taste all their own, a very different taste than adults, and if they are not using existing Web portals for children, the authors suggest that “it is either because they are unaware of their existence, or if aware of them, do not find them to be either visually or functionally attractive” (Large, Beheshti, Nnesset, & Bowler, 2004). Second, Large et al. (2004) found that children do not enjoy a meaningless spectacle of colors, pictures, and animations in their portal. They enjoy colorful interfaces only if the interfaces are functional for the information retrieval process and do not distract them from their search task. Information retrieval was found to be the major role of the portal in the eyes of the children. In relation to information retrieval facilities, children preferred the existence of keyword searching first, alphabetical browsing second, and subject categories browsing third. Another important design issue was found to be the search result display. Here children preferred “a brief but accurate summary of the document’s content in language that young people will understand” (Large, Beheshti, Nnesset, & Bowler, 2004).

As for working in an intergenerational team for designing a children’s Web portal, Large and his colleagues conclude that children and adults can be equal design partners only “to some extent” (Large, Beheshti, Nnesset, & Bowler, 2003). Adults are more often required to function as team leaders and guides; however, they must remain close in status to their younger team members and not take on the role of a teacher or other high status position (Large, Beheshti, Nnesset, & Bowler, 2004).
A similar research method was used by Bilal who, using a participatory approach, asked eleven 7th-grade students to draw their suggestions for a prototype interface of a search engine for children (Bilal, 2003). In this study, Bilal found four major components of the desired search engine from the children’s perspective. The first component is the existence of subject categories. In contrast with Large and his team’s findings, this study found the existence of hierarchical browsable subject categories to be very important for the young user, especially when related to school curriculum. Other components found to be of high importance include the existence of keyword and phrase search boxes alongside with the appropriate help and instructions, the existence of links to other Web search engines, and finally, a catchy and attractive name for the engine.

In a more recently published study, Bilal & Wang (2005) took the “children as designers” approach one step forward. While previous studies concentrated mainly on the visual interface design of the desired search tool, this study was the first to seriously address one of the more content-related aspects: the conceptual structure of the browsable subject categories in Web directories. Using a conceptual mapping technique, Bilal & Wang (2005) asked eleven middle school students to do two things: (a) sort and organize 17 science-related concepts (taken from two existing children’s Web directories, Yahooligans and KidsClick) from general to specific and (b) to draw two maps of the sorted concepts, one without links between the concepts and one with links between them. The authors later compared the maps, designed by the children, with each other as well as with the structures employed in Yahooligans and KidsClick. This study showed that children are able to sort science concepts into subject hierarchies. However, the sorting process was much easier for them, and it produced a higher similarity between the maps and structure employed by Yahooligans and KidsClick when the concepts were concrete (and not abstract) and when the concepts were familiar to the children from their school curriculum.

The Research

The purpose of this study is to investigate the process of categorization of information as perceived by elementary school children. Using the “children as designers” approach, we try to gain a comprehensive understanding of the structure and terminology of subject categories that children would like to encounter in Web directories specially designed for them.

Unlike Bilal & Wang (2005) and due to the finding that children spend much more time surfing the Web for fun than for educational purposes (Large, 2005), we decided to focus our investigation on leisure and entertainment subject categories appearing in Web directories.

In order to achieve our goal, we decided to use the card-sorting method of inquiry. Card sorting has been suggested as a simple yet effective method for understanding information and knowledge categorization abilities of children (Bilal & Wang, 2003, 2005; Borgman, Chignell, & Valdez, 1989), and even as a usability research technique for designing children’s computer products (Hanna et al., 1999). Card sorting is one of the techniques employed in user-centered design of Web sites (Katz-Haas, 1998) for children and adults alike. Envelopes were used to simulate subject categories. Revelle et al. (2002) used the enveloping method during the retrieval process. In the current study, the envelopes were used by the children in order to facilitate the creation of a hierarchical structure.

According to Piaget’s theory of cognitive development (Piaget, 1970; Bjorklund, 1995), between the ages of 7–11 (the so-called concrete operations stage), children are already capable of grouping objects or words into hierarchies of classes and subclasses. The classical experiment for measuring classification capibilities is the “class inclusion” problem, where the child is presented with eight pictures of animals, five dogs and three cats, and asked whether there are more dogs than animals. At this stage, they are capable of multiple classification as well, i.e., they have the ability to classify objects on more than one dimension such as color and size (see Siegler, 1991, p. 45). The categorization ability improves when (a) grouping concrete objects or words, rather than more abstract ones and (b) the children are more familiar with the concepts used in the tasks. It also improves as children get older (Bilal & Wang, 2003, 2005; Borgman, Chignell, & Valdez, 1989; Cooper, 2004). Categorization is viewed as one of the basic mental processes and the main way that people make sense of experience (Lakoff, 1987, p. ix). The information processing approach is a more recent cognitive development theory based on empirical findings. This theory can be viewed as a refinement of Piegatian theory (Hetherington & Parke, 1986, p. 420–421). Neimark, Slotnick, and Ulrich (1971) found that memory recall based on clustering is increasingly employed with age.

The Subject Cards

Subject cards for sorting were taken from existing Web directories. We started by extracting all subcategories of all hierarchy levels of the above-mentioned subjects from two existing Web site directories: Yahooligans and Walla (www.walla.co.il), an Israeli directory (written in Hebrew) originally intended for the use of adults but due to its seniority, scope, and publicity—and the fact that no analogous directory to Yahooligans is available for Hebrew speaking children—, it is commonly used by many Israeli children. Subcategories were extracted from four top hierarchy categories of the mentioned directories: (a) Arts & Entertainment and Sports & Recreation from Yahooligans and (b) Leisure & Entertainment and Sports from Walla.

After extracting all subcategories, duplicates were eliminated. We also eliminated all inappropriate subcategories from Walla (e.g., Sex, Gambling, Pub) and all subcategories from Yahooligans that were identified as meaningless for Israeli, Hebrew-speaking children (e.g., Christmas Movies).
More than 100 subcategory names (or later referred to as “terms”) remained, each of which was written on a 3” × 2” plastic-coated cardboard card. Cardboard and plastic were used due to their resistance and ease of use for children, who have been found to have difficulties with using regular slips of paper (Cooper, 2004).

During the pilot, we realized that we had too many cards covering too many topics, and some with very little importance or interest from the participants’ perspective. It was then decided to cut down the number of cards as well as to limit the scope to only five specific topics: Arts, Television, Music, Cinema, and Celebrities, with only 47 cards remaining. However, when we used the selected 47 cards in our pilot, it became very clear that children need concrete examples for abstract terms, and thus we added another 14 concrete cards such as names of actors, singers, TV shows, and musical instruments. In order to determine which concrete terms would be known to the children, we consulted one of the authors’ own child (6th grader). Eventually, we used 61 cards as can be seen in Table 1.

**The Participants**

Our study took place in Israel during the summer vacation (July–August) of 2004. Twelve groups of four children, either 4th- or 5th-grade graduates participated in the study. The children were chosen through convenience sampling, based on proximity or interest from the participants’ perspective. It was then decided to cut down the number of cards as well as to limit the scope to only five specific topics: Arts, Television, Music, Cinema, and Celebrities, with only 47 cards remaining. However, when we used the selected 47 cards in our pilot, it became very clear that children need concrete examples for abstract terms, and thus we added another 14 concrete cards such as names of actors, singers, TV shows, and musical instruments. In order to determine which concrete terms would be known to the children, we consulted one of the authors’ own child (6th grader). Eventually, we used 61 cards as can be seen in Table 1.

**TABLE 1.** An alphabetical list of all the cards.

<table>
<thead>
<tr>
<th>Action Movies</th>
<th>Actors</th>
<th>Ahava MeEver Lapina</th>
</tr>
</thead>
<tbody>
<tr>
<td>Artists &amp; Bands</td>
<td>Arts</td>
<td>Animation Movies</td>
</tr>
<tr>
<td>Cinema</td>
<td>Chiquititas &amp; Rebeles</td>
<td>Bands</td>
</tr>
<tr>
<td>Comedy Movies</td>
<td>Concerts &amp; Festivals</td>
<td>Classical Music</td>
</tr>
<tr>
<td>Drama Movies</td>
<td>Eyal Golan</td>
<td>Dag Nahash</td>
</tr>
<tr>
<td>Flute</td>
<td>Hamesh Vahetzi b</td>
<td>Erez Tafo</td>
</tr>
<tr>
<td>Keyboard Instruments</td>
<td>Leisure &amp; Entertainment</td>
<td>Harry Potter</td>
</tr>
<tr>
<td>Movie Production Studios</td>
<td>Movie Review</td>
<td>Lord of the Rings</td>
</tr>
<tr>
<td>MP3</td>
<td>Museums</td>
<td>Movies</td>
</tr>
<tr>
<td>Musicians</td>
<td>Nati Ravits d</td>
<td>Music</td>
</tr>
<tr>
<td>Painting</td>
<td>Pyjamas b</td>
<td>Naomi Shemer</td>
</tr>
<tr>
<td>Sculpting</td>
<td>Sarit Hadad</td>
<td>Record Companies</td>
</tr>
<tr>
<td>Sheshtoos b</td>
<td>Shrek</td>
<td>Shay Gabs b</td>
</tr>
<tr>
<td>Stringed Instruments</td>
<td>Telenovelas</td>
<td>Singers</td>
</tr>
<tr>
<td>Trumpet</td>
<td>TV</td>
<td>The Simpsons</td>
</tr>
<tr>
<td>Wind Instruments</td>
<td></td>
<td>TV programs</td>
</tr>
<tr>
<td></td>
<td></td>
<td>TV programs for kids</td>
</tr>
</tbody>
</table>

¹Hebrew TV shows, Ahava MeEver Lapina and Esti Hamechoeret, are actually Israeli Telenovelas; Cohav Nolad is the Israeli version of the American Idol contest. ²Hebrew TV shows for children. ³Israeli actors. ⁴Israeli singers (Naomi Shemer was a very famous Israeli composer as well as singer). ⁵Israeli bands. ⁶Israeli TV hosts. ⁷This word can be interpreted as both singing and poetry; all the children interpreted it as singing, while the Walla directory interpreted the word as poetry.
5th-graders was due to an earlier finding: Cooper (2004) found that younger elementary school children might not have the appropriate reading and sorting skills. Sixth-grade children were not included, because, in Israel, they are usually too occupied with the transition to middle school during their summer vacation.

The decision to work with groups of children instead of individuals (as was done by Bilal & Wang, 2003, 2005) was made during the pilot phase of the study, where we tested our method with three different combinations: two individual participants, two pairs, one group of three participants, and two groups of four, all of which were not included in the twelve groups studied. We realized that a group of participants has its clear advantages over an individual participant or a pair of participants, mainly in the area of indicative interaction. When a single child had to perform the assignment, he did it very quickly without always knowing why he did what he did. Because the element of understanding the way children construct subject hierarchies was our main goal, the "why" aspect was crucial; and in the case of an individual participant could have only been accomplished through the process of asking the child indicative questions about his work on each step. Questions of the type "why did you group these cards together" were unnatural to the children. When a second child participated, we realized that the dialog that emerged between the two participants can replace, to some extent, the questions asked by the researchers. In their attempts to agree upon the sorting process, which usually involved arguing, pairs of participants revealed very important insights about the crucial "why" aspect that are very indicative to us as researchers. The level of this indicative interaction increased dramatically when a third child was added, and it continued to grow when a fourth child entered the scene. Even though we never checked the possibility to work with more than four participants, we felt that very large groups might be too difficult to handle and thus not quite effective. Thus, we decided on four participant groups and asked them to agree upon each move in order to enhance their interaction. We also insisted that the children in each group were close friends or at least familiar with each other in a friendly manner because we wanted the children to interact freely among themselves without feelings of shyness, discomfort, or intimidation.

We had a single one-hour session with each of the twelve groups. Initially, we intended to meet the children for more than an hour; however, during our pilot we soon realized that children became quite impatient after the first hour or so (as we observed), perhaps due to the very hot climate of the Israeli summer. There were no indications, however, that their impatience was caused by frustration with the task itself, and the groups were able to finish the task within one hour.

The sessions took place, as mentioned before, on the living room floor of the “cooperative parents’” homes. We conducted the sessions in the late morning or early afternoon (as was most convenient for the children and their parents). Both of us attended all sessions together, one assuming an executive role (presenting the study to the children, giving instructions, clarifying vague communication, etc.), and the other documenting the session (even though we also recorded the session for later analysis) and clarifying children’s intentions regarding category structure. We kept no distance, sat with the children on the floor, and were free to ask the children questions whenever uncertainty about the children’s intentions, thoughts, or actions arose.

The Sessions

At the beginning of each session, we presented ourselves, our affiliation, and the purpose of the study. Because, as mentioned earlier, disclosing the true purpose of the study was found to bias results, at this stage of the session we explained only that our study aims to investigate the way children organize subject terms. The fact that the study’s implications are related to designing a Web directory was kept confidential until the end of the session. In one group (our last group for the summer), however, the cooperative mother did not follow our instructions and disclosed the real purpose to the children. This led the children to expect the use of computers and to be disappointed with the sorting process.

After the short introduction, each session started by placing all 61 cards in random order in front of the children. The children were given about five minutes to quietly observe the cards before any instruction was given. Then the children were asked to sort the cards into subject groups. They were not limited to a certain number of groups or number of cards within each group. However, in a few cases we saw that the children were a little confused by the logical connections between terms, and hence we decided to direct them to maintain relatively large and general groups at this stage. Children were also assured that the task is not a test and there were not any correct or wrong answers. Note that besides these initial instructions, we did not give any other specific instructions related to hierarchy creation. We made no remarks related to the categories, and interfered only occasionally when the discussion about the placement of a card reached a dead end. In these cases, we suggested that the children vote on the placement and pick the majority’s decision.

In addition, children were asked to put aside, in two separate piles, cards that they could not understand or were not familiar with and cards that did not belong to any of the subject groups they made. Children were also given a pack of blank cards and a marker, and they were instructed to use them in order to duplicate cards whenever they thought a card belonged to two or more different subject groups.

After completing this task, children were asked to name each group. They could either pick one existing card as the group’s name or write a name of their own on a blank card. Each group of cards was then picked up from the floor and organized in a pack, with the name card on top. In case the children disagreed on where to place a certain card or how to name a group, they were encouraged to discuss the problem and to try to reach an agreement, either by one member of the group convincing all the others or by a majority vote.
In the second stage of the session, we took each one of these relatively large groups composed earlier by the children, re-opened it in front of the children, and asked them to repeat the same process of sorting the cards in this group into subject subgroups and naming them. Through this process, we hoped to create a hierarchical structure. However, during the pilot, this was not very easy to accomplish. Children failed to recognize any hierarchical connections between the terms and thus very shallow hierarchies were formed. One possible reason for this could be that their vocabulary is too limited even for recognition, so that they can classify mostly concrete terms and are unable to create deep hierarchies. Another problem that we encountered at this stage was how to effectively simulate the normal view of a Web directory, where the user cannot see the categories under a given node. In order to encourage the creation of more complex hierarchical structures and hide the lower level categories, we decided to use colorful semi-transparent, plastic envelopes in five different sizes. Each of these envelopes represented one hierarchical level, and they were able to contain a number of smaller-sized envelopes as well as individual cards (see Figure 1).

As mentioned before, at this stage of the session, we separately re-opened each one of the relatively large subject groups created earlier by the children and asked them to repeat the same process of sorting it into subject subgroups and naming them. Using a piece of scotch tape, we first taped the card, representing the name of the top large group made earlier, on top of a big envelope. Then, we did the same thing with each of its subgroups: Each subgroup's name was taped on a smaller-sized envelope and all cards belonging to this subgroup were entered into this envelope. In a larger envelope, children could place either smaller envelopes (each having a name and containing some additional cards) or "non-enveloped" cards (in case they thought the card belonged to this category and not to one of the subgroups). These rules were established in order to simulate the actual structure of Web directories: Envelopes corresponded to subcategories and cards not taped to envelopes to content sites or to empty subcategories.

We repeated the same process of sorting, naming, and "enveloping" with every single subject group, subgroup, and sub-subgroup until the children were convinced there was nothing else to sort into subject groups. Children managed to complete this task in most cases. A few groups, however, lost their concentration during the re-sorting process and thus we had to limit the number of the re-sorted subgroups to the larger ones only.

Eventually, we received a number of big envelopes (identical to the number of primary subject groups children created at the beginning of the session) with the top-level names taped on top of them and each one containing all of its hierarchy structure represented by smaller and smaller envelopes. Each envelope contained smaller envelopes as well as individual cards that did not belong, according to the children, to any of the smaller envelopes (subgroups) of that particular envelope. During this process of reexamining the cards, some groups decided to slightly change the original groupings. The entire "enveloping" process is presented in Figure 1.

FIG. 1. The enveloping process.
At the final stage of the session, we tried to evaluate the effectiveness of the conceptual structure made earlier. In order to achieve this goal, we placed all top-level envelopes in front of the children and asked them to simulate a situation in which a classmate of theirs would step into the room and they would ask him to find one concrete term (a singer’s name, for example). We encouraged the children to think what steps that child would do, or more precisely, what envelopes that child would open in order to locate the desired card and if he would find it in the envelopes he would open. By this simulation technique, we encouraged the children to assess whether their sorting, naming, and “enveloping” process was suitable for the use of other children.

At the end of the session, we thanked the participants, and each child received a candy bar and a small wrapped present. (At the beginning of the summer, it was usually a toy, such as a yo-yo or a bubble making kit. As we came closer to the beginning of the school year, we gave more school-related presents such as notebook stickers and geometry kits.)

At this stage, we also revealed the actual purpose of the study and told the children that their insights would be later used in order to design a Web directory for the use of children. Participants were usually very thrilled to hear that they just contributed to the designing of a Web tool and in a few cases started to suggest other features that the tool should have.

**Measures**

As already mentioned, the purpose of this study was to understand the structure and terminology of subject categories that children would like to encounter in Web directories specially designed for them. In order to achieve this goal, we created a “consensus structure” that integrated the different structures created by the different groups.

In order to construct a “consensus structure,” we first calculated the maximum, minimum, average, and standard deviation values for each of the following descriptive measures: number of top-level categories, depth of structure, number of cards added, number of cards removed, and number of subcategories under the largest and smallest top-level category. Note that each card was viewed as a subcategory whether it was an abstract term like “TV programs for kids” or a more concrete one like a specific movie or a celebrity.

The “consensus structure” was created inductively. For each card, we examined how many times it appeared as a top-level category. If it appeared six times or more (for the majority of the groups), it became a top-level category in the “consensus structure” as well. Each of the remaining terms was examined, and if a term appeared six or more times under one of the top-level categories, it became a direct descendant of that top-level category. In case the six or more groups assigned the same term to more than one top-level category, it became a direct descendant of all the relevant top-level categories in the consensus structure. Terms that were never assigned six or more times in a certain position were not included in the consensus structure. If a term was excluded, then terms directly below it were reassigned to the parent node, and they were counted as direct descendants of the parent node in the structures created by the individual groups.

Qualitative descriptions of some of our observations also appear in the Findings and Discussion section. These comments and observations are an integral part of our findings.

**Validity and Reliability**

Validity and reliability in qualitative research are debatable concepts (see for example Golafshani, 2003). The rigor of the research design was ensured by taking several steps: (a) All the sessions were recorded, (b) notes were taken during the meetings, (c) both of us were present on all occasions, and (d) observations were summarized and compared immediately following each session. Twelve groups participated in the research, and the basic process was very similar in most groups. After the hierarchical structure was created, we asked the children to simulate a situation in which a classmate of theirs would step into the room and ask him to find one concrete term (triangulation).

**Findings and Discussion**

Generally, children were able to perform the given task. They did not have major difficulties with the sorting and “enveloping” process and mostly managed to complete the task. However, as mentioned earlier, three groups lost their concentration during the re-sorting stage, and thus we decided to re-sort only the larger categories in these groups. We assumed that impatience was mainly due to the very hot climate of the Israeli summer. All three sessions in which children got impatient took place in very hot rooms with no air conditioning. Another factor that might have affected impatience was the number of cards. Even though most groups managed to complete the task and none of them complained about the quantity, using a lesser amount of term cards might have been helpful to the three impatient groups.

Providing children with tangible entities, such as cards and envelopes, was found to be very useful. The envelopes simulated effectively the structure of a Web directory and children easily grasped the hierarchical subject structuring. As has been found in previous studies (Bilal & Wang, 2003, 2005; Borgman, Chignell, & Valdez, 1989; Cooper, 2004), we also found that the addition of concrete terms was very useful, both at the sorting stage and at the final stage of the session in which we asked the children to locate a hidden term from the perspective of another child. Children found it easier to sort and retrieve concrete terms rather than abstract ones; even though concrete terms were more often duplicated and assigned to several categories. The duplication process was quite controversial for some groups and caused lively discussions. We also found that understanding hierarchical structures improved with age. As was suggested by Cooper (2004), older children (in our case, 5th-grade graduates) were able to construct hierarchies more effectively. Two 5th-grade groups created the entire directory hierarchical...
structure (cards below cards) even before we introduced the envelopes (at the second part of the session). It is important to mention that both groups were from a relatively high socioeconomic background and studied in private schools (or semi-private, as is more accepted in Israel). Besides the understanding of hierarchical structures, we did not find any differences in task accomplishment and impatience levels between 4th- and 5th-graders in this study. We also did not find any other significant differences among genders and socioeconomic levels.

Our requirement that children in each group would be friends was found to be a very important factor, both in encouraging indicative interaction (as was explained earlier) and in facilitating communication among the group members and the researchers. Except for a little shyness in the first few minutes at the beginning of each session, children felt quite comfortable with our presence. Their friendship enabled them to freely discuss and argue almost without noticing the fact that they were being watched. The discussions and arguments over the task (encouraged by our requirement for consensus) were often loud and enthusiastic. Children also liked the fact that we sat with them on the living room floor. One child even remarked, “Wow, cool, you’re like us!” Parents, however, were more concerned with this fact and often offered us (the researchers) chairs.

Group dynamics was quite different among the groups. In some cases, a group leader emerged, while other groups were quite equitarian. We did not observe any connection between group dynamics style and age, gender, or socioeconomic level. We also did not observe any relationship between group leadership and the fact that one child (or two) was the “landlord” in each group. Group consensus upon every move was usually reached by majority voting (which obviously required 3 vs. 1 situation). This caused group members to try to persuade other members to agree upon different moves. Arguments and disagreements often arose among the children during the sessions. In most cases, children reached an agreement by themselves within a few minutes using persuasion techniques. On some rare occasions, researchers’ intervention was needed in order to achieve consent.

Consensus Directory Structure

Generally, there was reasonable agreement among the groups upon the desirable hierarchical structure of the terms used in this study. Out of the 61 terms given to the children, 52 (85.3%) were inserted into the “consensus structure.” The nine excluded terms are as follows: “MP3,” “Nati Ravits,” “record companies,” “artists & bands,” “concerts & festivals,” “dance,” “leisure & entertainment,” “Mozart,” and “museums.”

Among these terms, only the first three were omitted by the children during the sorting process. Five groups removed the term “MP3.” It was obvious that most of the children were not sure what “MP3” was, but the other groups were quite reluctant to admit this and insisted on inserting it in the sub-tree under music. Four groups removed the term “record companies,” not because it was unfamiliar to them but because they felt this subcategory was not an integral part of the structure. Three groups removed the actor “Nati Ravits,” probably because he was a less known actor. The children added only a few new cards during the sessions. We were not able to find any common characteristics among the cards added. All the groups duplicated at least one card. All cards duplicated were of concrete (and not abstract) terms. Sometimes the same card was duplicated more than once. The card of “Naomi Shemer,” for example, who was both a composer and a singer, was duplicated by all groups up to five times. The movies “Shrek,” “Lord of the Rings,” and “Harry Potter” were also duplicated a number of times (3, 2, and 3 times respectively), because the children felt that they belonged to more than one genre (among animation, drama, comedy, and action). In general, children tended to duplicate the cards containing people names in order to include them both under the appropriate subject node (e.g., singers, actors) and under the “celebrities” card.

Table 2 displays the descriptive measures calculated for the twelve groups.

Figure 2 depicts the consensus structure based on the structure produced by the individual groups. We deviated from the method for creating the structure for one category only: “cinema/movies.” We combined these two terms, whereas even though we define cinema as the place in which movies are projected, we saw that the children were confused about the relation between these two terms and used them interchangeably. They were also slightly puzzled about the difference between “TV” and “TV programs.” Contrary to our conceptualization of TV as the broader term that includes TV programs (as well as, for example, production companies, commercial TV, channels, and broadcasters), some of the groups felt that there was no need for both entries. The only category that appeared twice in the structure is “telenovelas” (soap operas), both under “TV” and under “TV programs.” Only a single group placed it in both places (the others placed it under either “TV” or “TV programs”), but this single double placement gave enough votes for the category to appear twice. This point also emphasized the confusion between “TV” and “TV programs.”

It is interesting to note the way some celebrities appear in the consensus structure: There were two cards for TV hosts, one of them was “Erez Tal” who hosts contests and talk shows

![Consensus Directory Structure](image)

### Table 2. The descriptive measures.

<table>
<thead>
<tr>
<th>Measure</th>
<th>Minimum</th>
<th>Maximum</th>
<th>Average</th>
<th>SD</th>
</tr>
</thead>
<tbody>
<tr>
<td>Number of top-level categories</td>
<td>3</td>
<td>8</td>
<td>5.17</td>
<td>1.27</td>
</tr>
<tr>
<td>Depth of structure</td>
<td>3</td>
<td>5</td>
<td>3.83</td>
<td>0.84</td>
</tr>
<tr>
<td>Number of cards added</td>
<td>0</td>
<td>3</td>
<td>0.92</td>
<td>1.08</td>
</tr>
<tr>
<td>Number of cards removed</td>
<td>0</td>
<td>4</td>
<td>1.92</td>
<td>1.24</td>
</tr>
<tr>
<td>Number of cards duplicated</td>
<td>1</td>
<td>5</td>
<td>2.92</td>
<td>1.08</td>
</tr>
<tr>
<td>Number of subcategories under smallest top-level category</td>
<td>1</td>
<td>11</td>
<td>4.08</td>
<td>3.18</td>
</tr>
<tr>
<td>Number of subcategories under largest top-level category</td>
<td>11</td>
<td>22</td>
<td>17.33</td>
<td>3.14</td>
</tr>
</tbody>
</table>
for adults. “Oded Menashe” is a TV host on the Children’s Channel. The first was categorized as a “celebrity,” while the last as an “actor.” Popular bands were not considered “celebrities.” Only one group placed them under celebrities; the rest of the groups positioned them under “music” or “artists & bands” (a category that does not appear in the consensus structure). The children view bands as entities and not as a group of musicians; hence, they do not see them as celebrities. Singers, on the other hand are famous people. Still, they were primarily associated with singing. However, two groups placed singers under “celebrities,” and a third group duplicated the cards for the individual singers and placed the cards both under “singing” and “celebrities.” An exception was “Naomi Shemer,” a famous composer who also performed her songs. She was placed a majority of times under “musicians,” five times in the “singing” subtree (either directly under “singing” or “singers”), and four times in the “celebrities” subtree. Her card was duplicated by four groups, with one of the groups duplicating the card twice. Actors were all viewed as celebrities, but Mozart was not viewed as a famous person except by a single group (in Hebrew, the term celebrities does not exist and the term accepted for celebrities is actually “famous people”). When in the final stage of each session, we asked the children to simulate how another child of their age would look for one of the cards of the singers in the envelopes, they often said that they would try to open the “celebrities” envelope and only as a second choice would they go to “music.” Thus, it seems that they are not quite confident about the placement of singers versus actors.

In addition, all movie titles appeared directly under movies/cinema in the “consensus structure.” However, “Harry Potter” and “Lord of the Rings” were placed four and five times respectively under “action movies” as well, and “Shrek” appeared five times under “animation movies.”

FIG. 2. The consensus structure.
The other terms that were excluded from the “consensus structure” are as follows: “artists & bands,” “concerts & festivals,” “dance,” “Mozart,” “museums,” and “leisure & entertainment.” Several groups felt that the card “artists & bands” was redundant. The card “concerts & festivals” was placed either under music or under “leisure & entertainment.” “Dance” was somewhat problematic. Some groups viewed it as part of arts and some as part of music. “Mozart” was either placed under “musicians” or under “classical music;” his name was duplicated only once. Several of the groups felt that “museums” is not an integral part of the structure, and were not sure where to place it.

The most problematic term of all was “leisure & entertainment.” In fact, in the existing Web directories from which we initially extracted the terms, “leisure & entertainment” was the top-level category. Our participants, however, failed to grasp the idea that “leisure & entertainment” was actually the common denominator of the entire pack of cards used in the study and thus the ultimate top-level category. Some of the groups were not sure about the meaning of “leisure & entertainment” to begin with. Some of them positioned it as a top category among other top categories (but not as the ultimate top-level category), while others under “TV,” “movies,” or “music.” Only two groups (both girls, 5th grade) noticed the relationship between “leisure & entertainment” and the rest of the cards. One grasped it immediately, while the other was still entangled with it at the beginning: “We should duplicate leisure and entertainment
to every group here, because in television you have leisure and entertainment, and in movies you have leisure and entertainment, and also in music and in everything.” However, after a while a girl commented: “Hey, wait a minute, if we can put leisure and entertainment in every group, maybe we can put it on top, maybe every group is connected to it.” Finally, the entire group agreed to this idea.

As mentioned earlier, in the final stage of each session we asked the children to evaluate the category structure by simulating the moves that would be taken by another child in order to retrieve specific concrete terms. Over all, children positively evaluated the structures they made earlier; they found their structures to be logical and easy to navigate. The only further suggestion made by the children at this stage was that it might have been helpful to include all persons’ names under “celebrities” as well as under their subject categories.

Obviously, a more proper evaluation of the structure made could be achieved mainly by children who did not participate in the study. We intend to achieve this goal in a future study.

Comparing our “consensus structure” with the structures of existing Web directories, especially the ones we used for term extraction in this study, shows that there are only slight differences between the “consensus structure” and the basic structures we used. Figure 3 shows the relevant parts of the Walla directory and Yahooligans! as they existed in the summer of 2005. We added some of the concrete examples (names of singers, TV shows, etc.) to the cards and they were not part of the original structure. Comparing the two figures, one can see that the original structure is slightly more complex.

Conclusions

In this exploratory research, groups of 4th- and 5th-grade children were asked to create a directory structure on popular topics. A list of terms was presented to them and they were asked to base their structure on these terms. They were also allowed to suggest their own terminology, to discard some of the suggested terms, and to duplicate terms.

The children were all at the end of the concrete operations stage (Piaget, 1970) and, according to Piaget’s cognitive theory, they were supposed to be able to create hierarchical structures. At this age, the theory predicts that it is still easier for them to work with concrete examples. Our findings support the theory regarding concrete objects: once more, concrete terms (names of musical instruments, movies, TV programs, and famous persons) were included; the task became easier for them. However, at first we were puzzled by their lack of ability to create meaningful hierarchical structures. Piaget based his finding about the classification capabilities of children between the ages 7 and 11, on the class inclusion problem, which shows that the child understands the concept of subgroups. This, however, is not the same as the ability of creating multi-level hierarchies, with occasional crosslinks (duplicated terms). We were able to overcome the cognitive obstacle by introducing the “enveloping” method. A possible reason for the success with this method is that the categories and subcategories became more concrete and tangible than without the use of the envelopes, and they enabled the children to grasp the concept that when viewing a title of the category, the contents under the category is hidden.

Overall the children succeeded in creating coherent structures and there was reasonable agreement between the groups on the final structure. The consensus structure we created based on their structures includes more than 85% of the original terms.

The children duplicated, discarded, and created a relatively small number of cards, which indicate that they were more or less satisfied with the suggested terminology. Only a few problems with the terminology were located—the most remarkable one was the ambiguous relation between cinema and movies and between “leisure & entertainment” and the rest of the subject cards. This finding is somewhat similar to the finding by Bilal and Wang (2003, 2005), where most of the children placed the term “medicine” under “hospitals” and not the other way around as it appears in the KidsClick directory.

We strongly recommend children to be included in the design process of the content side of the development of Web search tools for children, just like Druin and Hourcade (2005) who emphasize the need for “design methods with children.” The findings of this study show the importance of consulting the future users (the children in our case) in designing the structure of a site or a directory. Currently, we are working on developing measures that will allow us to compare the consensus structure with the individual structures; and in the future we intend to add content (Web sites) to the structure, again with the help of children.

References
