Available online at www.sciencedirect.com
ScienceDirect
Information Processing and Management 44 (2008) 910–930
INFORMATION PROCESSING & MANAGEMENT
www.elsevier.com/locate/infoproman

Symbolic links in the Open Directory Project

Saverio Perugini *

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

Received 29 May 2006; received in revised form 26 June 2007; accepted 29 June 2007
Available online 17 August 2007

Abstract

We present a study to develop an improved understanding of symbolic links in web directories. A *symbolic link* is a hyperlink which makes a directed connection from a webpage along one path through a directory to a page along another path. While symbolic links are ubiquitous in web directories such as Yahoo!, they are under-studied and, as a result, their uses are poorly understood. A cursory analysis of symbolic links reveals multiple uses: to provide navigational shortcuts deeper into a directory, backlinks to more general categories, and multiclassification. We investigated these uses in the Open Directory Project (ODP), the largest, most comprehensive, and most widely distributed human-compiled taxonomy of links to websites, which makes extensive use of symbolic links. The results reveal that while symbolic links in ODP are used primarily for multiclassification, only few multiclassification links actually span top- and second-level categories. This indicates that most symbolic links in ODP are used to create multiclassification between topics which are nested more than two levels deep and suggests that there may be multiple uses of multiclassification links. We also situate symbolic links *vis à vis* other semantic and structural link types from hypermedia. We anticipate that the results and relationships identified and discussed in this paper will provide a foundation for (1) users for understanding the usages of symbolic links in a directory, (2) designers to employ symbolic links more effectively when building and maintaining directories and for crafting user interfaces to them, and (3) information retrieval researchers for further study of symbolic links in web directories.

Keywords: Cross-references; Hypermedia; Information access; Information hierarchies; Link analysis; Link typing; Navigational search; Open Directory Project; Symbolic links; Taxonomic navigation; Web directories; World Wide Web

1. Introduction

The problem addressed by this paper is improving the understanding of symbolic links in web directories. Informally, a symbolic link is a hyperlink which makes a directed connection from a webpage along one path through a website to a page along another path. The concept of a symbolic link in general is not unique to electronic environments. Cross-references, such as a 'see also stylists' annotation in the the 'beauty salons' section of the *Yellow Pages*, function similar to symbolic links. In UNIX, soft or symbolic links – those with an `l`

* Tel.: +1 937 229 4079; fax: +1 937 229 2193.
E-mail address: saverio@udayton.edu
URL: http://homepages.udayton.edu~perugisa.

0306-4573/$ - see front matter
doi:10.1016/j.ipm.2007.06.005

to the left of the file permission mode, e.g., `lrwx- - - - - - -` – are used for similar purposes (Marsden & Cairns, 2003).

Symbolic links are most common in large web directories such as Yahoo! Such directories, which are organized along a hierarchy of categories (e.g, Business, News, or Sports), are the web's analog to the *Yellow Pages*. Web directories represent one of three major paradigms of searching the web and, as a result, serve as a gateway or *starting point* to web resources for many users (Baeza-Yates & Ribeiro-Neto, 1999). Moreover, while 'the web coverage provided by directories is very low (less than 1% of all webpages), the answers returned to the user are usually much more relevant' as such directories often exclude low quality sites (Baeza-Yates & Ribeiro-Neto, 1999). Users interact with these directories by progressively drilling down the categories to find pointers to websites of interest (called *structure guided browsing*; Baeza-Yates & Ribeiro-Neto, 1999). The underlying goal of symbolic links is to improve information access. However, designers of web directories use symbolic links in multiple ways, introduced below, within the scope of improving information access.

1.1. Uses of symbolic links in web directories

We use the sample web directory shown in Fig. 1 to illustrate the following uses of symbolic links. We call every page in a web directory a *topic* page. We use the words *topic* and *category* interchangeably in this article. Topic pages are divided into non-leaves and leaves. We call a page in a directory with at least one link to another topic page in the directory a *non-leaf*. In Fig. 1, pages 1–7 are non-leaves. Conversely, we call a page in a directory with links only to external webpages (i.e., those which are not part of the directory) a *leaf*. In Fig. 1, pages 8–14 are leaves. We also say that a *path* is an ordered set of hyperlink labels from the root page of a directory to a leaf. For instance, ≺computers: software: music≻ and ≺arts: theatre: vocal@: jazz≻ are paths in Fig. 1. A *hard path* (consisting of only hard, or non-symbolic, links) does not contain a symbolic link. Therefore, above, the former path is a hard path while the latter is not. We use the term *sequence* to refer to a subset of a path, e.g., ≺computers: software≻ is a sequence in the former path above.

- **Shortcuts** (Scs): One use of a symbolic link is to provide a navigational shortcut to a page nested deeper in a directory. For example, in Fig. 1, the symbolic link labeled 'memory@' provides a shortcut from webpage 3 to 14 bypassing the intermediate page 7. Such a usage gets users to the information they desire in fewer steps. We define a *shortcut* as a symbolic link whose target can be reached via a path, without symbolic links, through its source. Notice that a shortcut need not end in a *leaf*.

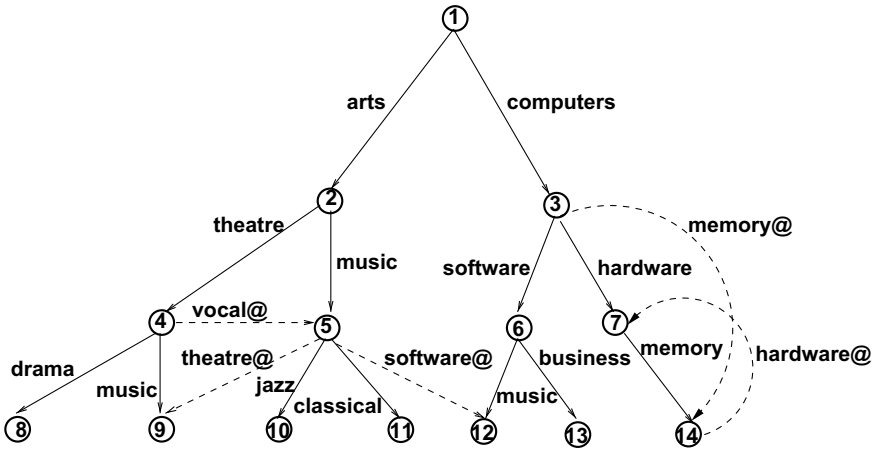

Fig. 1. Sample web directory, simplified for purposes of presentation, with characteristics similar to those in Yahoo! Nodes correspond to webpages and directed edges correspond to hyperlinks between pages. Symbolic links are indicated by dashed edges and hyperlink labels ending with @.