Link Clouds and User-/Community-Driven Dynamic Interlinking of Resources

Andreas Nauerz and Martin Welsch
IBM Research and Development
71032 Böblingen, Germany
{andreas.nauerz|martin.welsch}
@de.ibm.com

Fedor Bakalov and Birgitta König-Ries
University Jena
07743 Jena, Germany
{fedor.bakalov|koenig}
@informatik.uni-jena.de

Abstract

During the last years we have observed a shift in the way how content is added to web-based systems. Earlier, dedicated authors were responsible for adding content, today entire communities contribute. As a consequence these systems grow quickly and uncoordinated. New ways had to be found to organize and structure content.

Tagging has become one of the most popular techniques to allow users (and entire user communities) to perform this structuring autonomously. But, not only because current tagging systems have their flipsides (e.g. synonyms and polysems lead to littered tag spaces making it difficult for users to find relevant content), we argue that tagging is sometimes an abstraction layer not necessarily needed. In many scenarios users just want to interlink content fragments (resources) with each other. In this paper we present an approach allowing users, i.e. the community, to collaboratively define relations between arbitrary content fragments. They can interlink any source with any target. We allow for personal interlinking of resources as well as collaborative interlinking. In the latter case we visualize, for each single resource, available interlinks in what we call link clouds, a concept comparable to tag clouds. We finally leverage the knowledge about the interlinks between resources' for building personal (or community) navigation structures and for performing content recommendations.

The concepts presented are being prototypically implemented within IBM's WebSphere Portal and can be presented in a live demo at the workshop.

1 Introduction

Today, web-based systems are often comprised of a huge amount of content. They are no longer exclusively maintained by IT departments, instead, Web 2.0 techniques are used increasingly, allowing user generated content to be added. These systems grow quickly and in a more uncoordinated way as different users possess different knowledge and expertise and obey to different mental models.

The continuous growth makes access to really relevant information difficult. Users need to find task- and role-specific information quickly, but face information overload and often feel *lost in hyperspace*. Thus, users often miss out on resources that are potentially relevant to their tasks, simply because they never come across them. On the one hand, users obtain too much information that is not relevant to their current task, on the other hand, it becomes cumbersome to find the right information and they do not obtain all the information that would be relevant.

As users (and entire communities) have been enabled to contribute content, mechanisms have been introduced to allow categorizing, organizing and structuring this content, too. Particularly tagging and rating, which have become very popular collaboration techniques, provide new means doing this kind of categorization. It can add valuable meta information and even lightweight semantics to web resources. Tagging allows non-expert users to develop folksonomies that categorize content available in the system.

In our previous work [Nauerz *et al.*, 2008] we developed several tagging engines, which e.g. allowed arbitrary annotators, e.g. human users or analysis components (for automated tagging), to annotate any resources. The analysis of users tagging behavior allowed us to model their interests and as well as semantic relations between resources, and thus to perform reasonable recommendations and adaptations. In [Nauerz *et al.*, 2009] we also introduced new tagging paradigms like alien tagging, reputation-based tagging, quantitative tagging, anti tagging, tag expiry, contextual tagging, and described how these can be used to refine our models and to perform even more valuable adaptations or to issue more valuable recommendations.

But tagging engines also have their flipsides, though: synonyms and polysems lead to littered tag spaces making it difficult for users to find relevant content. Users suffer from retrieving content actually not being of interest or, vice versa, from not retrieving content that actually would be of interest when exploring the tag space. Even worse, tagging requires users to invest work and thus time: they need to come up with proper tags and assign them to the appropriate resources. If users just want to interlink resources with each other this is an unnecessary overhead, probably one reason why in most tagging systems not more than approximately 20% of all users tag content (cp [Al-Khalifa and Davis, 2007] and [Sen *et al.*, 2006]).

In this paper we present an new approach for solving the problems just mentioned. We argue that if we allow people to contribute content, we should also allow them to organize and structure this content leveraging their collective wisdom. But we want to enable them to do so without forcing them to come up with proper tags for resources. We want to relief them from this overhead if not really necessary.

We regard tagging as an interesting approach to catego-

rize content and see dynamic interlinking as an interesting accompanying approach to relate content fragments to each other.

Thus we present an approach allowing users, i.e. the community, to collaboratively define relations between arbitrary content fragments. They can interlink any source with any target. We allow for personal interlinking of resources as well as collaborative interlinking. In the latter case we visualize, for each single resource, available interlinks in what we call link clouds, a concept comparable to tag clouds. We finally leverage the knowledge about resources' interlinking for building personal navigation structures and for performing content recommendation.

2 Related Work

As already indicated, a lot of newer approaches to allow users to categorize, organize and structure content autonomous have been made by introducing abstraction layers like tagging.

But only few work has been done to find solutions allowing users to directly interlink resources. Even lesser work has been invested to find solutions leveraging knowledge about the interlinks created to aggregate link clouds (cp. 3.3), to construct personal- or create new navigation menus (cp. 3.4), or link flows (cp. 3.6), or to do content recommendations (cp. 3.5).

So far, most approaches rely on means to automatically improve link structures. Adapting link structures (including link sorting, link annotation, and link hiding as well as generating links) based on user profiles etc. has been performed a lot, approaches are e.g. described in [Brusilovsky, 1996]. Even earlier work on computed and adaptive linking included the implicit linking mechanism described by [DeRose, 1989], as well as the work described in [Bieber and Kimbrough, 1992] and [Stotts and Furuta, 1991].

Other early approaches to automatically interlink resources focus on computing links based on relationships or similarities between texts or passages of text, where a link is not defined as a pointer from one hypertext node to another, but rather as a query that leads to a different node. [Allan, 1996] describes how documents can be analyzed and automatically interlinked if similar. [Bodner and Chignell,] describes how text analysis can be performed on what they refer to as source nodes and target nodes. Depending on the similarity of both nodes links are automatically generated between those. A similar approach is described in [Wilkinson and Smeaton,] which is also based on the determination of the relationships between nodes to interlink them.

In [Nauerz and Welsch, 2007] we have described our approaches for automatically adapting link structures (and navigation topologies).

Some approaches to manually create links between resources (e.g. documents) are described in [Carr *et al.*, 1998]. But even with these approaches, where the community is given the power to decide which resource to interlink to which other resource, concepts like the ones mentioned above have not been pursued.

3 Concepts

Web-based systems are comprised of content fragments (also referred to as resources). These resources can be structuring elements like web pages, or with respect to Portal systems also pages and portlets. These resources provide users with content and services. On a more fine-granular basis resources can be any identifiable information unit, an image, a video, a document, a text passage, and so forth. Different resources provide different information, which can still be related. E.g., there might be pages part of an Enterprise Information Portal that provide means to book flights, hotels, cars or trains - different pages with similar use cases.

Prior to the Web 2.0 era these resources have been brought into relation by some central instances, usually administrators or content authors. However, those superimposed structures were not necessarily compliant to users' mental models and therefore resulted in significant effort to find the information needed. This became even worse, once user generated content was added, where the structure did not follow the design the administrator had in mind. fig. 1 shows the structure of a sample system: four branches exist below the root node. Along the first branch authors have put everything having something to do with "flying", e.g. pages that provide information about airports (location, arrival and departure times of flights, etc.), travel regulations (official regulations and internal company regulations), and finally a page to book a flight. Along the second branch authors have put everything have something to do with "hotels", e.g. pages that provide information about hotels at different locations as well as a page to finally book a hotel. Along the third branch authors have put similar pages having something to do with "cars and trains". Underneath the fourth branch users find pages to do their travel expense.

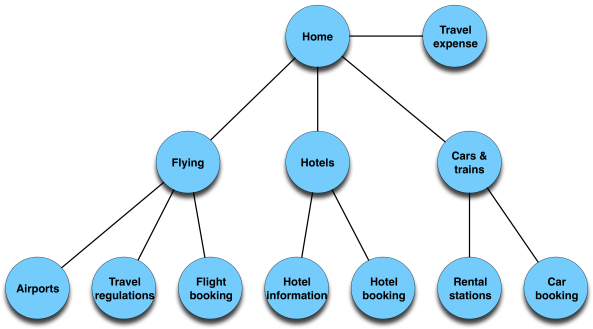

Figure 1: Structure of the sample system

Experienced users know about their favorite airports, the external and internal travel regulations and so forth - they just want to do their bookings. Given the structure above users would have to perform a lot of (unnecessary) clicks to traverse the *booking* pages. With the availability of a tagging engine users could have had tagged (which would have been work, too) these pages with the term "booking", but even then users would have to fire up the tag cloud, select the right tag, analyze the result list of resources that have been tagged with the selected tag and select the right one out of those being presented. Moreover, there could be more pages tagged with the term "booking" but being irrelevant in the scenario described.

As said, the question is, why, if users contribute content, we do not allow them to organize and structure this content, too? Or, in other words, why we do not allow them to interlink resources independently from what administrators or content authors thought is the correct structure?

3.1 Personal Dynamic Interlinking

Private personal interlinks behave similar as private tags as described in [Nauerz *et al.*, 2009]. They can only be seen by the user who created them. fig. 2 exemplarily visualizes how private personal interlinks can be created within a typical Web Portal solution. First, the user navigates to the resource where he wants to interlink from, the source resource. A resource can be a page, a portlet or anything uniquely referencable. There he clicks a button which triggers the linking process. Next, he navigates to the resource he wants to interlink to, the target resource and clicks a button which finishs the linking process and establishes the interlink between both resources.

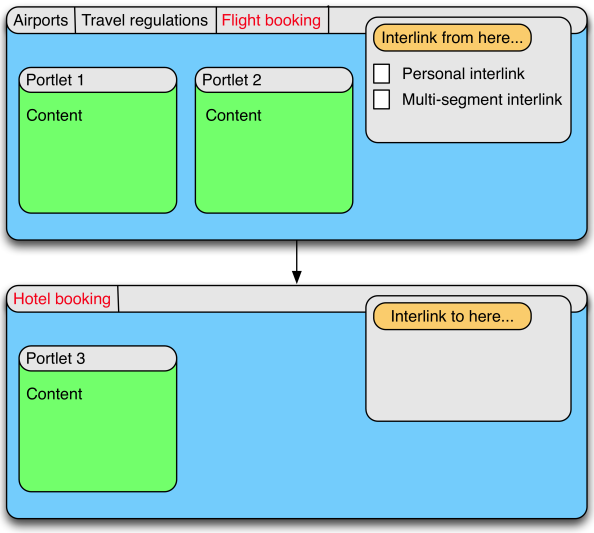

Figure 2: Creating interlinks

This way every user can interlink the resources he personally thinks should be related, totally independent from what an administrator or content author had in mind who always try to create structures satisfying majorities but not necessarily single users. He can manually create personal shortcuts and cross-references between related content. This way navigating through the system can be personalized and speed-up.

In the sample described earlier a user might be one of those experienced users that usually want to do his booking just by sequentially traversing the three *booking* pages and the *travel expense* page. Thus he would create three personal interlinks as depicted in fig. 3 (red connectors), one from the *flight booking* page to the *hotel booking* page, one from the *hotel booking* page to the *car booking* page, and one from the *car booking* page to the *travel expense* page. Next time he is doing his bookings he can follow this path by doing three clicks only, just following his personal interlinks.

3.2 Collaborative Dynamic Interlinking

The real power and benefits of dynamic interlinking becomes evident when allowing collaborative dynamic interlinking. A collaborative dynamic interlink created by one user can be seen by all other users, too. Creating collaborative interlinks is done similar as creating private interlinks, except that a checkbox indicating that the next interlink to be created should be a private interlink has not to be selected (cp. fig. 2).

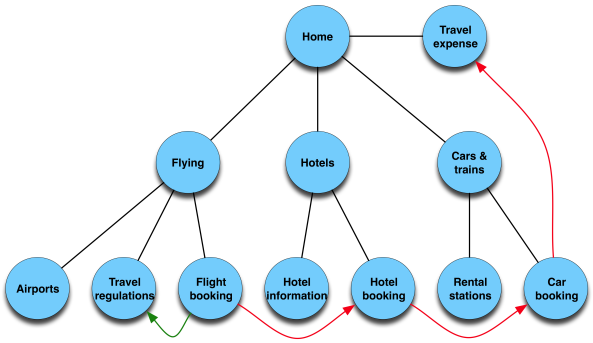

Figure 3: Personal interlinks

The private interlinks created before could have been created as collaborative interlinks, too. Thus, one or more experienced users could have set interlinks between the *booking* pages. This can help people that want to do their booking for the very first time. These users do not need to search for one *booking* page after the other anymore, instead they can follow the interlinks available.

Private and collaborative interlinks can be mixed, of course. E.g., in addition to the collaborative interlinks (red connectors) interlinking the *booking* pages, less experienced users might want to interlink from these pages to the corresponding pages providing information for travel regulations (green connectors) (cp. fig. 3 again).

3.3 Visualizing Dynamic Interlinks

An important aspect is that of course multiple interlinks can be created from any resource to any other; similar each single resource can be the endpoint of several interlinks pointing to it. Depending on users needs interlinks could also be added between other *booking* pages (cp. fig. 4 where solid as well as dotted red lines represent collaborative interlinks), e.g. between the *flight booking* page and the *car booking page*, the *flight booking* page and the *travel expense* page and so forth. This could be done by users that e.g. do never need all three booking pages, e.g. because they never book a car and want to skip the corresponding page.

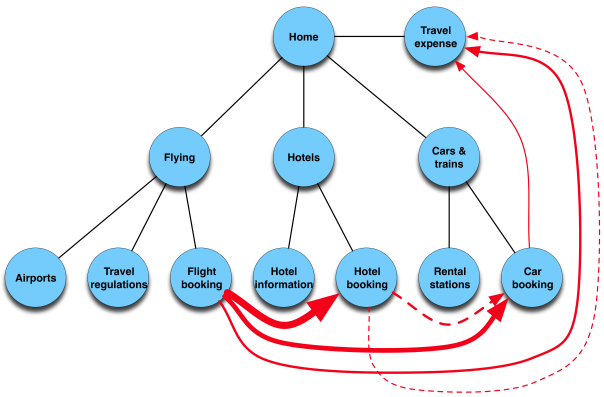

Figure 4: Collaborative interlinks

This is similar to what we observe in collaborative tagging environments, where single resources can be tagged with multiple tags. The most important concept to visualize tags (and their importance) for single resources or a

set of resources are tag clouds. Tag clouds display which tags are available and how often these have been applied (with respect to what one is looking at); more often applied tags are regarded as more important tags which are usually presented in a larger font size.

We propose a similar mechanism, which we refer to as **link clouds**, for visualizing dynamic interlinks. When navigating from one resource to another, the user-/community created personal interlinks can be used in addition to the links that were originally created by an administrator or content author.

So, e.g. if collaborative interlinks have been created as described above, when being on the *flight booking* page there could be a link to the superior *flights* page as modeled by the content author. But additionally a link cloud could display interlinks to the *hotel* page, the *cars* page and the *travel expense* page. If most users navigate from the *hotel* page to the *cars* page more people would interlink these two pages. So, if e.g. 10 users interlink the *flight booking* page to the *hotel* booking page, 5 from the *flight booking* page to the *cars* booking page and 1 from the *flight booking* page to the *travel expense* page the first linkage would be regarded the most important one, the second one the second most important one and so forth. fig. 4 visualizes this as the thicker solid red connectors represent interlinks set by more users.

Link clouds visualize this importance to the users. Different solutions can be thought of. In one embodiment (cp. fig. 5) link clouds could look like tag clouds presenting a description of the target resource they are linking to. Depending on the importance of the available collaborative interlinks (derived by how often a certain interlink has been set) some targets could be presented more prominent (larger font size) than others.

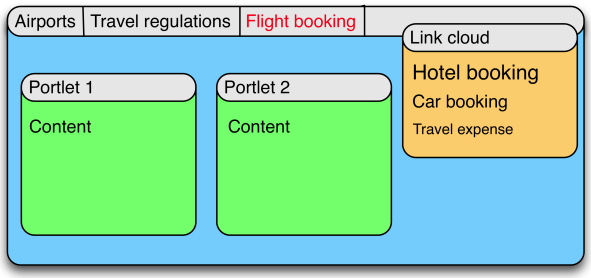

Figure 5: Link clouds

3.4 Personal and Community Navigation

Taking into consideration all personal and collaborative interlinks available in the system we can provide users with additional navigation menus, accompanying the one originally created by an administrator or content author, from which they can select. Therefore we provide users with a pull down menu, displayed at the top corner of every page, where he can select between these navigation menus:

- Original navigation
- Personal navigation
- Community navigation
- Aggregated navigation

The *original navigation* represents the navigation as created by an administrator or content author not containing any personal or community interlinks; The *personal navigation* adds personal interlinks to the *original navigation* so that these can be used from within the standard navigation menu. The *community navigation* adds collaborative interlinks to the *original navigation* and the *aggregated navigation* adds personal and collaborative interlinks to the *original navigation*.

It is also possible to display a navigation menu comprised of personal or community interlinks only, not containing the *original navigation* at all. This can be controlled via an additional check-box.

With respect to our previous sample, fig. 6 shows the *aggregated navigation*, which contains all administrator created links, as well as all collaborative interlinks and the user's personal interlinks. The same figure without the green connectors would represent the *community navigation*, without the red connectors the *personal navigation* and without the green and red connectors the *orgininal navigation*.

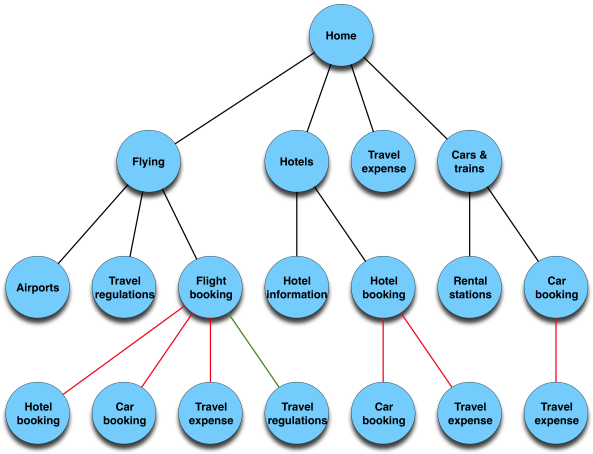

Figure 6: Personal-/community navigation

3.5 Content Recommendation

Leveraging the knowledge about incoming and outgoing dynamic interlinks for any resource allows us to perform related content recommendations. Three scenarios can be thought of (cp. fig. 7):

Forward linking (red lines in fig. 7) describes the most trivial case. With respect to our previous sample we might know that interlinks exist pointing from the *flight booking* page to the *hotel booking* page, *car booking* page and *travel expense* page. Thus we know that all these three target pages have something to do with the source page and can be recommended when being on the source page.

Backward linking (red lines in fig. 7) describes the second case. We might know that the *flight booking* page, the *hotel booking* page, and the *car booking* page link to the *travel expense* page. Thus we could recommend these three source pages when being on the target page.

Sideward linking (red lines in fig. 7 again) describes the third case. Again, we might know that interlinks exist pointing from the *flight booking* page to the *hotel booking* page, *car booking* page and *travel expense* page. Thus, there might not only be a relationship between the source and target pages, but also among the sources (or targets) themselves. Thus, a user being on the *hotel booking* page

might also be interested in the *car booking* page as both are referenced from the same source page.

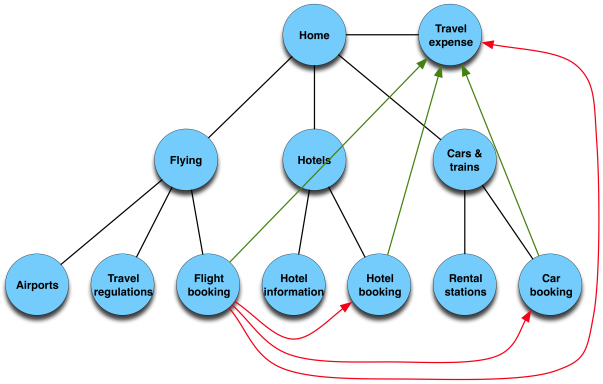

Figure 7: Forward, backward and sideward linking

3.6 Multi-Segment Interlinking

We also allow for doing more than just interlinking one resource to exactly one other. We refer to a continuous sequence of interlinks as **link flows**.

Such paths could be manually created by users (an additional check-box in the UI, cp. fig. 2, allows to do so), or, in a more sophisticated variant detected and recorded by the system. Latter could be based on following, from one resource to another, the "top" interlink (the one set by most users), or on analyzing which available interlinks users follow, again from one resource to another, most often.

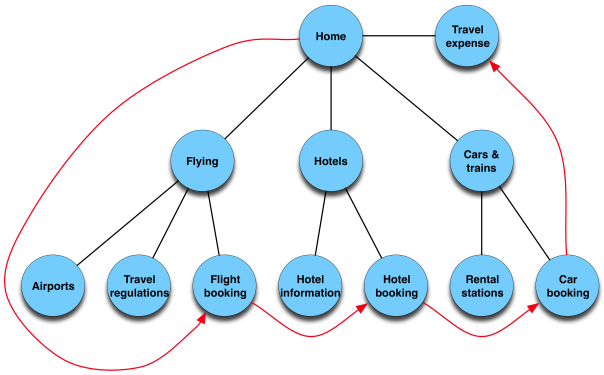

Figure 8: Multi-segment linking

With respect to our sample such a path could be comprised of the resources *home* page, *flight booking* page, *hotel booking* page, *car booking* page, and *travel expense* page (cp. fig. 8).

4 Conclusion and Future Work

In this paper we have presented a new approach to interlink arbitrary fragments of web elements by defining relations between them. Users can do this both only personally, but also in a collaborative way. Such relations can then be visualized and used for navigation purposes. For individual use such personal interlinks work like personal tags. Users can create their own optimized linking/relationship network independent from schemes defines by administrators or content authors. In the collaborative case the created link patterns are visible to all and provide added value by reflecting common or strongly used relations. We also proposed to eploit the explicitly generated knowledge about interlinking for content recommendation. Furthermore interlinking can be extended beyond single steps thus creating link flows or paths that could e.g. encompass an entire task to be performed. Users can always freely select between the original or the new enhanced link cloud navigation. These concepts have been prototypically implemented in IBM's WebSphere Portal.

In the future, we intend to work on visualization and UI related extensions. We are looking forward to discuss these ideas and get feedback. Each aspect still needs to be evaulated in terms of both usabilitiy and usefulness.

IBM and WebSphere are trademarks of International Business Machines Corporation in the United States, other countries or both. Other company, product and service names may be trademarks or service marks of others.

References

[Al-Khalifa and Davis, 2007] Hend S. Al-Khalifa and Hugh C. Davis. Towards better understanding of folksonomic patterns. In *Proc. of the 18th Conf. on Hypertext and hypermedia*, pages 163–166, New York, NY, USA, 2007. ACM Press.

[Allan, 1996] James Allan. Automatic hypertext link typing. In *HYPERTEXT '96: Proceedings of the the seventh ACM conference on Hypertext*, pages 42–52, New York, NY, USA, 1996. ACM.

[Bieber and Kimbrough, 1992] Michael P. Bieber and Steven O. Kimbrough. On generalizing the concept of hypertext. *MIS Q.*, 16(1):77–93, 1992.

[Bodner and Chignell,] Richard Bodner and Mark Chignell. Dynamic hypertext: querying and linking. *ACM Comput. Surv.*, page 15.

[Brusilovsky, 1996] Peter Brusilovsky. Methods and techniques of adaptive hypermedia. *User Modeling and User-Adapted Interaction*, 6(2-3):87–129, 1996.

[Carr *et al.*, 1998] L. A. Carr, D. De Roure, W. Hall, and G. Hill. Implementing an open link service for the world wide web. *World Wide Web*, 1(2):61–71, 1998.

[DeRose, 1989] S. J. DeRose. Expanding the notion of links. In *HYPERTEXT '89: Proceedings of the second annual ACM conference on Hypertext*, pages 249–257, New York, NY, USA, 1989. ACM.

[Nauerz and Welsch, 2007] Andreas Nauerz and Martin Welsch. (Context)Adaptive Navigation in Web Portals. In *Proc. of the Intl. IADIS WWW/Internet Conference*, Vila Real, Portugal, October 2007.

[Nauerz *et al.*, 2008] Andreas Nauerz, Stefan Pietschmann, and Rene Pietzsch. Social recommendation and adaptation in web portals. In *Proceedings of the International Workshop on Adaptation for the Social Web*, Hannover, Germany, 2008.

[Nauerz *et al.*, 2009] Andreas Nauerz, Fedor Bakalov, Martin Welsch, and Birgitta Knig-Ries. New tagging paradigms for content recommendation in web 2.0 portals. In *Proceedings of the International Workshop on Adaptation and Personalization for Web 2.0 (in conjunction with the 1st and 17th International Conference on User Modeling, Adaptation and Personalization)*, Trento, Italy, 2009.

[Sen *et al.*, 2006] Shilad Sen, Shyong K. Lam, Al Mamunur Rashid, Dan Cosley, Dan Frankowski, Jeremy Osterhouse, F. Maxwell Harper, and John Riedl. tagging, communities, vocabulary, evolution. In *Proc. of the 20th Conf. on Computer supported cooperative work*, pages 181–190, New York, NY, USA, 2006. ACM Press.

[Stotts and Furuta, 1991] P. David Stotts and Richard Furuta. Dynamic adaptation of hypertext structure. In *HYPERTEXT '91: Proceedings of the third annual ACM conference on Hypertext*, pages 219–231, New York, NY, USA, 1991. ACM.

[Wilkinson and Smeaton,] Ross Wilkinson and Alan F. Smeaton. Automatic link generation. *ACM Comput. Surv.*, page 27.