Social and Behavioral Aspects of a Tag-based Recommender System

Frederico Durao
IWIS — Intelligent Web and Information Systems
Aalborg University, Computer Science Department
Selma Lagerlöfs Vej 300, DK-9220 Aalborg-East, Denmark
Email: fred@cs.aau.dk

Peter Dolog
IWIS — Intelligent Web and Information Systems
Aalborg University, Computer Science Department
Selma Lagerlöfs Vej 300, DK-9220 Aalborg-East, Denmark
Email:dolog@cs.aau.dk

***Abstract*—Collaborative tagging has emerged as a useful means to organize and share resources on the Web. Recommender systems have been utilized tags for identifying similar resources and generate personalized recommendations. In this paper, we analyze social and behavioral aspects of a tag-based recommender system which suggests similar Web pages based on the similarity of their tags. Tagging behavior and language anomalies in tagging activities are some aspects examined from an experiment involving 38 people from 12 countries.**

I. Introduction

Collaborative tagging systems have become increasingly popular for sharing and organizing Web resources, leading to a huge amount of user generated metadata. Tags in social bookmarking systems such as *del.ici.ous*[1] are usually assigned to conceptualize, categorize, or sharing a resource on the Web so that users can be reminded of them later and find their bookmarks in an easy way. Invariably, tags represent some sort of affinity between user and a resource on the Web. In this sense, tags from social bookmarking systems represent a potential mean for generating personalized recommendations. Following this assumption, our previous work [4] introduced a personalized tag-based recommender approach which suggests similar Web pages based on the similarity of their tags. In that opportunity, we only evaluated the user's satisfaction about the received recommendations. In this paper, we extend the previous work by reexamining our (prior) results from the social and behavioral point of view. In particular, we assess aspects such as tagging user behavior, diversity and cultural background of the participants and language issues. The goal of this study is therefore to identify how social and behavioral aspects interfered in our previous results in order to improve our prior approach from the learned lessons. In addition, we believe that the findings of this work may be enjoyed by other recommender systems which tackle experiments in an open and multicultural environment. The contribution of this paper is therefore an analysis of social and behavioral aspects of results from a tag-based recommender system.

The paper is organized as follows: In Section 2 we discuss related work. Section 3 revisits the previous work and presents new findings. Section 4 addresses behavioral aspects of our previous experiment while Section 5 focus on social aspects. Section 6 presents the conclusion and future works.

[1]http://delicious.com

II. Related Work

Tags have been recently studied in the context of recommender systems due to various aspects. Social and behavioral issues are aspects constantly addressed in the evolutionary process of tag-based recommender systems. Golder and Huberman [5] address perhaps the most significant formal study of collaborative tagging system. They analyze how tags assigned to an individual resource (in the case of del.ic.ious, Web resources) change - or more specifically, stabilize - over time. Additionally, Golder and Huberman examines the tag roles (or intentions) when assigned to Web resources. [9] addresses user experience as a social factor to build their Web recommender system. For means of personalization, they utilize folksonomy tags to classify Web pages and to express user's preferences. By clustering folksonomy tags, they adjust the abstraction level of user's preferences to the appropriate level. Another study in the direction of tagging behavior can be seen in [10], which proposes a model for tagging evolution based on community influence and personal tendency. It shows how four different options to display tags affect user's tagging behavior. [2] studies how the tags are used for search purposes. It confirms that the tags can represent different purpose such as topic, self reference, and so on and that the distribution of usage between the purposes varies across the domains. Other works such as [11] and [8] coined the term *emergent semantics* as the semantics which emerge in communities as social agreement on tag's meaning based on its more frequent usage instead of the contract given by ontologies from ontology engineering point of view.

In this paper we share the social and behavioral concerning (closely related to the ones addressed in this section) in the terms on how they interfered in the overall result of our previous work.

III. Contextualizing: The Previous Work and New Findings

In our previous work [4], we proposed a tag-based recommender system which suggests similar Web pages based on the similarity of their tags. The proposed approach extends basic

978-0-7695-3872-3/09 $26.00 © 2009 IEEE
DOI 10.1109/ISDA.2009.19

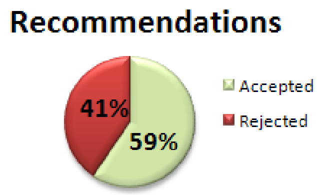

Fig. 1. Acceptance and rejection rate

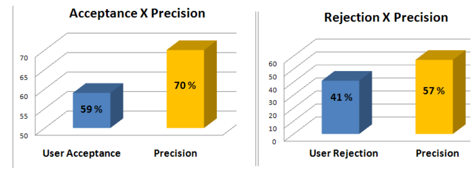

Fig. 2. Comparing accetance and rejection against precision

similarity calculus of the tags with external factors such as *tag popularity*, *tag representativeness* and the *affinity between user and tag*. In a nutshell, we utilize a *cosine similarity* [6] measure between tag vectors to calculate basic similarity of resources. *Tag popularity* is measured as a count of occurrences of a certain tag in the total amount of Web pages. The term frequency measure is used to compute *tag representativeness* for a certain resource. The *affinity between a user and a tag* is calculated as a count of how many times the user utilized the tag at different Web pages. We propose a model which considers all these four factors in a normalized way and gives a ranking of Web pages for particular user. As priory mentioned, our approach was evaluated in an experiment with 38 participants from 12 countries interested in different subjects. The goal was to measure the degree of satisfaction of users about the received recommendations. Before taking part of the experiment, the participants had an introduction about the experiment and instructions on how to proceed towards the recommendations. The goal of this introduction was to clarify our expections about their participations on the experiment. In brief, we created a *del.icio.us* user account for each participant on which he/she was invited to create at least 10 bookmarks with minimally 3 tags each. One month later, the participants were invited to answer a questionnaire with the generated recommendations besides other questions about tagging activity and their participation in the experiment (our previous work only evaluated the generated recommendations). Data for our experiment was collected from *del.icio.us* in November 2008 comprising 5542 tags and 1143 bookmarks. More details about our previous work can be seen at [4].

Figure 1 shows the overall user's satisfaction rate in which **59%** of whole recommendations succeeded and **41%** of them was rejected. Regarding cultural diversity of people and consequently data in our experiment, further data analysis was undertaken aiming at exploring new findings out of the experiment. Additionally to the data analysis, a questionnaire was sent to the participant asking questions about tagging behavior and the overall participation in the experiment.

A. Analyzing the Accepted and Rejected Recommendations

An interesting analysis was to compare the user's satisfaction (acceptance and rejection) result against the precision of our recommendations. This could reveal reasons why correctly generated recommendations were rejected and wrongly generated recommendations were accepted. We consider a recommendation as *correct* if it contains one or more tags similar to the tags used by the receiver for describing his/her bookmarks. The user's satisfaction results were taken from our previous results as shown in Figure 1 while the precision was achieved comparing the tags assigned to the accepted (or rejected) recommended items against each (set of) participant's tags. Recall was not considered because the only the top 5 recommendations were taken into account in the experiment. More about precision metric can be found at [1].

Figure 2 shows the user's satisfaction (acceptance and rejection) rates just followed by the calculated precision rates. Based on the results obtained, it is observed a difference of **11%** between the accepted items and the precision and a significant difference of **16%** between the rejected items and the precision.

A deeper analysis into the set of recommendations which corresponds to the **11%**, showed that such rejected items were probably known by the receivers. In particular, we had a case in which a user rejected a recommended item about *reuters rss* because he had bookmarked the *reuters main page*. From this situation, two possible argumentations are reasonable: i) the user who rejected the item had previous knowledge about it or ii) he/she did not know what it was about. Another finding was the fact that our participants use to reject very popular Web sites such as *Google* and *Facebook*. As people are mostly interested in new stuffs, popular Web sites increase the chance of being rejected. Further analysis into the set of recommendations which corresponds to the **16%**, showed that such items were accepted because some participants were strict to the experiment, i.e. they did not judged the received recommendations in accordance with the instructions.

According to the instructions of the experiment, the participants were inquired to choose recommendations similar to their bookmarks taking into account the tags they have used. Therefore, any judgement out of this rationale means that the participants are not following the experiment strictly.: For those participants who did not follow the instructions properly (both sets 11% and 16%), we understand they accepted or rejected items analyzing other factors such as novelty and interestingness. Nevertheless, we do not discard such behavior because this is a natural reaction of humans. Inevitably people get attracted by things they do not know but want to experiment. This input however becomes a potential point of interest for our future studies. Thus, it is important to state that this work is not proposing any formal model for *interestingness (or novelty) of recommendations*. Instead, we are only collecting evidences from the experiment which shows

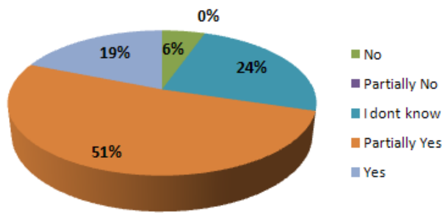

Fig. 3. Rates for interesting recommendations

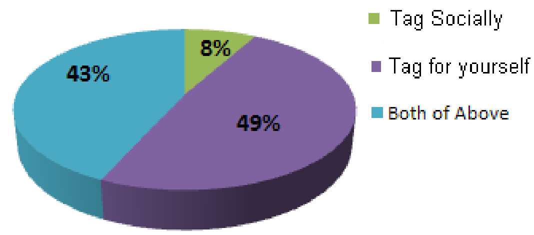

Fig. 4. Tagging Behavior

that some participants found interesting recommendations.

B. Evidences of Interesting Recommendations

We asked our participants if they had identified interesting recommendations, even for those which were rejected. The accurate meaning of "*interestingness*" addressed in this question is precisely understood as "*the power of attracting or holding one's attention because it is unusual, exciting*". Based on the results achieved (as seen in Figure 3), **66%** of our participants responded *yes* or *partially yes* which is a significant finding regarding the fact that **41%** of our recommendations was rejected. As we have pointed out before, the goal of our previous work was simply to recommend items which share similar tags, but not interesting things. Unexpectedly we surprised users with recommendations they (probably) had no idea about but enjoyed.

The collection of such evidences pushes our future work to design a formal model for interesting recommendations. We believe that this type of recommendations keep users committed to system and discover additional content beyond the currently viewed page. However, we also have in mind that being *interesting* does not necessarily mean **efficient** since such recommendations can distract users from the application context. In terms of personalization, this issue sounds as an opportunity for discovering more information about what users like or dislike. The key point for our future work is therefore to discover what makes a recommendation *interesting* and when it should be applied.

IV. Behavioral Aspects

In this section, we focus on issues related to tagging activity and user behavior while tagging.

A. Tagging Behavior: How do Users Tag?

In general, people tag either to organize resources or due to social motivations [7]. In order to identify these practices, we asked how the participants usually tag: *socially*, *for themselves* or *both*. Tagging socially means that they have used terms (tags) that, in general, anyone understands and agrees. Tagging for themselves means they usually use self reference tags or particular vocabulary for quick remembering. Figure 4 shows that **8%** of the participants tag exclusively socially, **49%** tags exclusively for themselves and **43%** tags for both reasons. From the results obtained, we are happy to evidence that our participants have strong social spirit. Nevertheless, for computational purposes, to be "social" means *to make use of* "popular" terms so that it contributes significantly for finding similarities between tags. This reasoning motivated us to assess the popularity of each tag of ours participants. From the computational perspective, we consider a tag "popular" if it provides a list of most related tags which is given by the most frequent *tag pair* co-occurrence among different resources. We represent a *tag pair* as $Tp(t_i, t_{ii})$ where t_i and t_{ii} are distinct tags of a set of tags T. Focusing on a *tag pair* $Tp(t_i, t_{ii})$, which is assigned to a set of resources R, all other *tag pairs* $tp \in Tp$ assigned to resources $r \in R$ are considered as co-occurring *tag pairs*, whereas the frequency of the co-occurrence depends on how often the *tag pairs* has been assigned by users. The more often a *tag pair* $Tp(t_i, t_{ii})$ is co-assigned, the higher is the chance of t_{ii} be part of the most related terms of t_i.

Given the fact that this statistical approach performs well with a huge corpus, we then crawled *tag pairs* from *del.ici.ous*. The motivation for using *del.ici.ous* is the fact that it is a big repository of tag available on the Web and our set of tags is quite limited to provide accurate degree of relatedness between two tags. In this context, we analyzed the top T most related terms of our participant's tags and end up with the following outcome: **43%** of the tags provided related terms which means that they are *popular* whereas **57%** of the whole set of tags did not generate related terms which lead us to classify them as *unpopular*. This balanced result, in fact, confirmed what the participants have expressed in the questionnaire: they tag socially and for themselves. As expected, the particular behavior of tagging socially impacted positively in the success of our recommendations since it is easier to find similar bookmarks which contain popular tags. An excerpt of unpopular tags is: *zipf, splcompanies, erasmus, mellon, recommnder, pthreads, campinas, danahboyd, antipiracy, kis, nielsen, futuretelling, spfc, hotelbooking, innebandy, pirateglossary, ectel08 human_computer_interaction, sports_&_outdoors, prolearn*.

B. Tag Intention and Purposes

We also evaluated the intention of the participants while tagging their bookmarks. We asked the participants to specify the purpose behind of the tags they were using.

Results from the questionnaire shows (as seen in Figure 5) that **38%** of the participants placed tags with the *intention of identifying what (or who) is it about* and **25%** tagged with

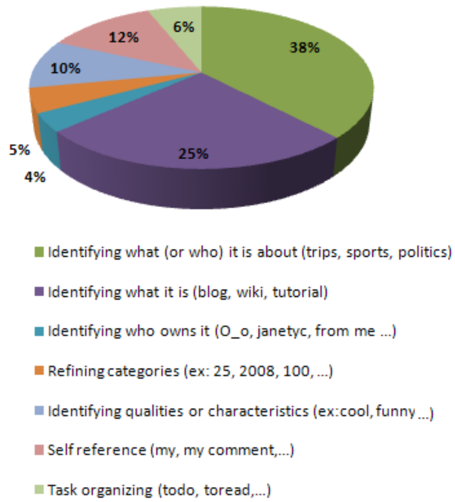

Fig. 5. Outcome of Tag Intentions

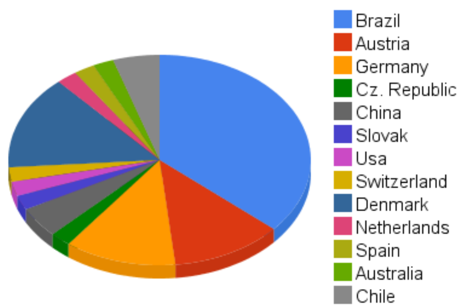

Fig. 6. Countries from where our participants come from

the purpose of *identifying what it is*. Focusing on these most representative results, we conclude that the major intention of our participants was to *conceptualize* the tagged resources. Self reference tags also deserve special attention (**12%**) because it shows that users tend to bookmark personal contributions on the Web that they own, have authored or have contributed. We believe that the *"my"tags* (or equivalent sense) provides potential hints for personalization since they explicitly say what the user does, owns or like.

Additionally to the provided tag-intention alternatives (see Figure 5), some users externalized other purposes while tagging, for instance, *to provide a translation* - some participants tagged simultaneously in German and English. This finding motivated us to investigate language anomalies and how they had affected our acceptance results.

C. Language Anomalies

Regarding the cultural diversity of our participants (as seen in Figure 6), we looked for language issues which may have impacted our recommendations. According to Figure 7, **84.2%** of the participants received recommendations in English while **15.8%** received recommendations in other language.

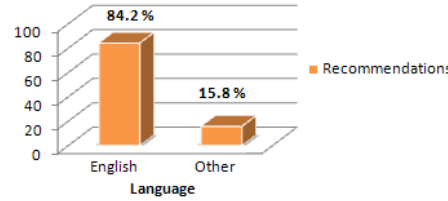

Fig. 7. Final results

Regarding English as a universal language, we did not consider an issue when a non native English speaker receives recommendations in English. However, recommendations in any another language sent to non native speaker of such language could be a problem. In this sense, we investigated those 15% in which language anomalies were detected and we found out interesting cases:

- Recommended items in Danish were accepted by Brazilians. Further investigation revealed that such Brazilians are living in Denmark and have used Danish tags to organize their Danish daily acitivities.
- A particular recommended item whose title was written in Portuguese, it was tagged in English. Although the recommendation was properly generated, it was rejected. Further we identified the receiver was not Brazilian and likely could not read articles written in Portuguese.
- Some participants classified their bookmarks by tagging the language of the resource, i.e. "dutch_page", "german", "pt". This evidence only occurred for bookmarks whose language was not English. We outline this special sort of annotation as an interesting issue for information retrieval purposes.

In conclusion, we understand that language was an issue which contributed for some rejections. In a multicultural context like we evidenced in our experiment, language treatment is highly necessary to increase the overall acceptance rate.

V. Social Aspects

In this section, we focus on issues related to knowledge about others tag-based recommender system and identification of groups by interests.

A. Knowledge about Tag-based Recommendater System

We asked the participants whether they have experienced others tag-based recommender systems or similar mechanisms which utilize tags to generate recommendations.

According to the results (as seen in Figure 8), only **27%** of the participants had experienced similar systems, **67.6%** have never heard about it and **5.4%** have no idea about what it is. The results showed that tag-based recommender systems or similar mechanisms are not too popular as we expected among our participants. This result could be sufficiently normal if **73.7%** of our participants were not from computer science field as seen in Figure 9. From this result, we may speculate that, in general, people believe that tag application is limited

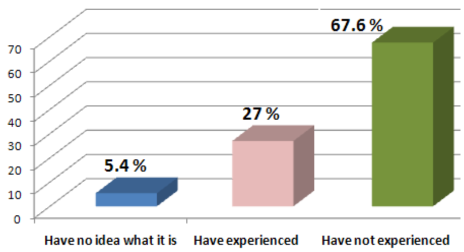

Fig. 8. Knowledge about tag-based recommendater system

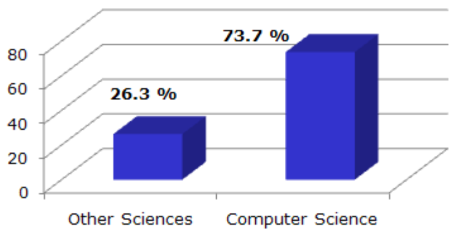

Fig. 9. Areas of participants

to annotate and describe resources on the Web. We face this outcome as an opportunity to keep our research on the track and put more effort for future works.

B. *Grouping Users by Interests*

In this section we investigate the relationship between users and tags. Since we have a notion about the areas of our participants (as shown in Figure 9), we investigated whether the participants are (predominantly) using tags related to their area of interest in their bookmarks. We then grouped users by their most frequent tags (MFT) because we believe this information may quickly reveal the user's interests instead of analyzing the whole set of tags.

The calculation of MFT is not a difficult task, however, to define which of them best represent the user's interest depends on more precise analysis. The *amount of tags a user has* and the *frequency on which each tag appears* are variables which should be taken into account. For this approach, we selected those tags whose tag frequency is **70%** closer to the top (most) frequent tag. The number *70* was chosen because values above of it could not allow us to find the MFT of 80% of our participants. In addition, values bellow of it would not precisely identify the MFT of our users.

Once the MFT was calculated, tree main groups were identified, as seen in Figure 10.

- **Technology** - In this group, there are participants which have utilized tags related to specialized technologies such as "unix, semantic Web, design, hci".
- **Information and Web Entertainment**: In this group there are participants which have utilized tags related to broad vehicles of communication, Web applications where end users look for social networking and Web entertainment such as "cnn, twitter, blog, facebook".

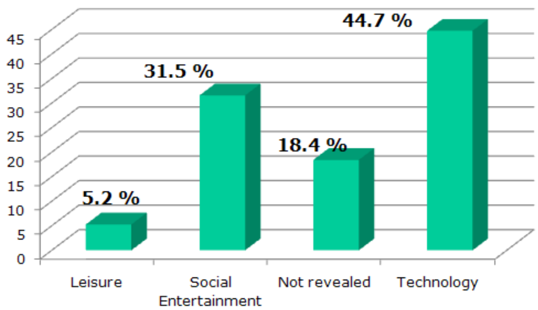

Fig. 10. Groups emerged from the Most Frequent Tags (MFT)

- **Leisure** - In this group there are participants which have utilized tags related to fun or recreational time such as "football, sport, cinema, movie, trip".
- **Not Revealed** - In this group are all participants in which we could not infer a precise interest.

For clustering users in groups, the K-Means algorithm [3] was aplied over the MFT of all participants. Since the overall amount of MFT was not high, we identified four clusters (*k=4*) as the most representative number for grouping users into the mentioned categories. However, for social network purposes, a more granular categorization should be applied considering that each identified category encloses other sub-categories. However, for this experiment such categories are reasonably sufficient to assess whether the participants are tagging resources on the Web related to their area of actuation. Thus, analyzing the results expressed in Figure 9 against the outcome shown in from Figure 10, we realized that **94%** of the participants which are in the group of *Technology* and **66%** of the participants which are in the group of *Information and Web Entertainment* belongs to the *Computer Science* area. This outcome allows us to safely state that ours participants are actually using tags related to their area of actuation. A special attention must be made to the fact that **18.4%** of all participants could not be classified in any one of those groups. We interpret this result as a limitation of MFT method to radically infer user's interest.

VI. Conclusion and Future Works

This paper analyzed social and behavioral aspects of a tag-based recommender system which suggests similar Web pages based on the similarity of their tags. We analyzed possible reasons why correctly generated recommendations were rejected and wrongly generated recommendations were accepted. We collected evidences that some rejected recommendations were considered interesting by our participants. We further realized that our participants have great social spirit while tagging, although they tag for themselves as well. We realized that the major intention of our participants was to *conceptualize* the tagged resources. We identified that some language anomalies contributed for increasing the rejection rate of our recommendations. We realized that although most of our participants come from Computer Science area, they are not familiar with tag-based recommender systems. Finally, we identified that ours participants are using tags (for bookmarking resources on the Web) related to their area of actuation.

All findings listed above are inputs for our future works. In this sense, we aim at creating a formal model to describe interesting recommendations besides analyzing in which circumstances this kind of recommendations are advised. We also intend to investigate how our recommender system can benefit from self-reference tags in order to find similar interests among users. The role of tags will also be analyzed since identical tags may have different intentions when assigned to resources. Finally, we intend to find ways to overcome language anomalies and build (or reuse) existing user models which assist predicting user's preferences to perform more accurate recommendations.

VII. Acknowledgment

The research leading to these results is part of the project "KiWi - Knowledge in a Wiki" and has received funding from the European Communitys Seventh Framework Program (FP7/2007-2013) under grant agreement No. 211932.

References

[1] Ricardo Baeza-Yates and Berthier Ribeiro-Neto. *Modern Information Retrieval*. Addison Wesley, May 1999.

[2] Kerstin Bischoff, Claudiu S. Firan, Wolfgang Nejdl, and Raluca Paiu. Can all tags be used for search? In James G. Shanahan et al., editor, *Proc. of the 17th ACM Conference on Information and Knowledge Management, CIKM 2008*, pages 193–202, Napa Valley, California, USA, October 2008. ACM.

[3] Richard O. Duda and Peter E. Hart. *Pattern classification and scene analysis*. John Wiley & Sons, 1973.

[4] Frederico Durao and Peter Dolog. A personalized tag-based recommendation in social web systems. In *International Workshop on Adaptation and Personalization for Web 2.0*, 2009.

[5] Scott Golder and Bernardo A. Huberman. The structure of collaborative tagging systems. *Journal of Information Science*, 32(2):198208, April 2006.

[6] David A. Grossman and Ophir Frieder. *Information Retrieval: Algorithms and Heuristics (The Information Retrieval Series)(2nd Edition)*. Springer, 2nd edition, December 2004.

[7] Cameron Marlow, Mor Naaman, Danah Boyd, and Marc Davis. Position Paper, Tagging, Taxonomy, Flickr, Article, ToRead. In *Collaborative Web Tagging Workshop at WWW2006, Edinburgh, Scotland*, May 2006.

[8] Peter Mika. Ontologies are us: A unified model of social networks and semantics. *Web Semant.*, 5(1):5–15, 2007.

[9] Satoshi Niwa, Takuo Doi, and Shinichi Honiden. Web page recommender system based on folksonomy mining for itng '06 submissions. In *ITNG '06: Proc. of the Third Intl. Conference on Information Technology: New Generations*, pages 388–393, Washington, DC, USA, 2006. IEEE.

[10] 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 Pamela J. Hinds and David Martin, editors, *Proc. of the 2006 ACM Conference on Computer Supported Cooperative Work, CSCW 2006*, pages 181–190, Banff, Canada, November 2006. ACM.

[11] Steffen Staab. Emergent semantics. *IEEE Intelligent Systems*, 17(1):78–86, 2002.