

# **An Adaptive Systems Approach to the Implementation and Evaluation of Digital Library Recommendation Systems.**

**JOHAN BOLLEN and LUIS M. ROCHA**

Modeling, Algorithms, and Informatics Group (CCS-3)

Los Alamos National Laboratory, MS B256

Los Alamos, New Mexico 87545, USA

e-mail: rocha@lanl.gov or rocha@santafe.edu

## ***Citation:***

Bollen, Johan, Luis M. Rocha [2000]. "An Adaptive Systems Approach to the Implementation and Evaluation of Digital Library Recommendation Systems." In: *Research and Advanced Technology for Digital Libraries: 4th European Conference, ECDL 2000*. Lectures Notes in Computer Science, Springer-Verlag, pp.356-359.

# An Adaptive Systems Approach to the Implementation and Evaluation of Digital Library Recommendation Systems.

Johan Bollen and Luis M. Rocha

Los Alamos National Laboratory  
Computer Research and Applications Group  
Los Alamos, NM (USA)

**Abstract.** The focus for information retrieval systems in digital libraries has shifted from passive repositories of information to recommendation systems that actively participate in retrieving useful information, and can furthermore learn from the retrieval behavior of users. We propose a novel evaluation methodology for such systems based on the concepts of shared knowledge structures, and system development reliability and validity.

## 1 Evaluation of Recommendation systems

### 1.1 Precision and Recall for traditional IR

Traditionally, IR evaluation methodologies have focused on the assessment of the set of documents retrieved or recommended to users [2]. Performance evaluation of IR systems, both in experimental [3] and analytical [4] approaches, is based on measures that use the sets of retrieved and relevant documents:  $RETR$  and  $REL$ . The most common measures are  $recall = \frac{RETR \cap REL}{REL}$  and  $precision = \frac{RETR \cap REL}{RETR}$ . The problem lies in determining the elements of the relevant set  $REL$ , which ultimately depends on the subjective judgments of human experts.

Traditional IR performance evaluations avoid the inherent subjectivity of relevance evaluations, even when techniques such as cross-validation, expert sampling (see e.g. the TREC (Text REtrieval Conference) standard [3]) and query expansion techniques [5] are used. They can therefore not take into account the requirements of specific communities of users.

### 1.2 Evaluation of Recommendation Systems

Interactive, adaptive RSs, contrary to traditional IR systems, actively recommend information items and adapt the relational information for the set of documents and keywords they operate on to their user communities' characteristics. For these kinds of systems we are mostly interested in the conditions that allow the associations among documents, keywords, and between the two, to reflect the knowledge of their particular community of users: *validity*, and if and how the adaptive algorithms of RS converge to the knowledge of users: *reliability*. We therefore propose a simulation approach that enables us to assess how recommendation systems interactively generate representations of the knowledge of their community of users, rather than attempting to describe the semantics or relevance of individual documents, as current IR methodologies do.

### 1.3 Validity and Reliability with Shared Knowledge Structures

**Collective choice and shared knowledge structures** We have taken an approach in which we operationalize the aggregation of user knowledge as shared knowledge structures (SKS). From the comparison of the recommendation system's knowledge base and the SKS, we can then device stability and validity performance measures of the recommendation system's adaptive behavior. The work of Richards [6] on collective choice has been particularly relevant in this context. Richards' model is based on the assumption that a population of agents needs to make a collective choice on a set  $\mathcal{A}$  of alternatives  $a_i$  on which a shared knowledge structure (SKS) is imposed. The SKS will be represented by a non-directed, connected graph  $W(A)$ . This graph is understood as the particular associative knowledge structure that all agents share in an environment. Agents are assumed to prefer one of the alternatives  $a_i$ , thus, for each agent, there is a preference partial order  $D_i(a)$ . Rather than modeling the capacity of a collection of agents to reach a single choice, we are interested in modeling an adaptive recommendation system's ability to reliably adapt to a valid representation of the patterns of associations in the agents' SKS.

**A simulation methodology** We propose the following simulation approach to a performance evaluation of an adaptive recommendation system. Let  $K$  and  $D$  be, respectively, a set of keywords and documents that are derived from a number of given database records and combined into a large set of alternatives  $\mathcal{A}$ . A particular community of users utilizes a subset of  $A \subset \mathcal{A}$ . The SKS of this community of users then is  $W(A)$ . To model such a community, a population of agents is generated with partial orders  $D_i(A)$  starting on a particular element  $a_i \in A$ . Because  $A$  is typically very large, we use incomplete partial orders  $D_{i,p}$ , that is a partial order starting at  $a_i$ , but extending only to the nearest levels  $p$ . Parameter  $p$  indicates how much of the SKS a given agent "knows" about.

This community of agents is then set to interact with a given recommendation system which will aggregate the partial knowledge of all agents into the adaptive structure of the information resource. We start with an initial graph  $V(A)$  representing the initial associative structure of the information resource we wish to model.<sup>1</sup> We then let the population of agents that share a SKS  $W(A)$  interact with  $V(A)$ , using any number of adaptation algorithms. In this simulation methodology, validity can be defined as the similarity between the final structure of  $V(A)$  (after completion of a sufficient number of iterations of the adaptation algorithm) and  $W(A)$ . The reliability measure requires a number of in parallel developed  $V(A)$ s, to determine whether the adaptive algorithm converges on one and the same associative structure in different runs, given the same population of agents.

## 2 The Identification of User communities

The above described methodology relies critically on our ability to generate a population of agents that share a SKS. SKSs for a number of communities the Los Alamos National Laboratory's (LANL) Research Library (<http://lib-www.lanl.gov/>) have been generated based on keyword and document semantic proximities [7] to demonstrate the viability of the proposed methodology.

<sup>1</sup> This can be an initially random structure for information resources that have no initial set of document links such as citations.

## 2.1 User Log Clusters

The Research Library at the LANL is a networked digital library, i.e a large part of its repository has been digitalized and can be downloaded from the library's web site. Extensive web logs keep track of users' retrieval patterns. A similarity matrix for journal titles had previously been generated based on the co-occurrence of journal titles in user retrieval paths in the February 1999 Research Library web logs (for a more detailed description of this technique, see [1]). A hierarchical cluster analysis was performed on this matrix, revealing a number of persistent journal clusters. Three clusters were selected for further analysis based on their size and content (table. 1): *Information Sciences* (cluster1), *Molecular Biology* (cluster2) and *Non-Linear Science* (cluster3).

cluster 1	cluster 2	cluster3
<b>Inform. Sci.</b>	<b>Mol. Biol</b>	<b>Non-Lin. Sci</b>
Inform. Proc. Letters	Analyt. Chim	Prev. Med.
Inform. Sciences	Tetrahedr. A	Chaos
	J. Mol. Biol.	Siam J. Comp
	Tetrahedron	Quant. Res.
	Tetrahedron L.	Siam Rev.
		Arch. Env. C.

**Table 1.** Three selected LANL RL journal clusters derived from the library's web log co-occurrences

A user community was derived by determining a set of users that had frequently downloaded articles published in the journals in the journal clusters. For each of those communities a list of the 20 most frequently articles and their associated keywords was compiled <sup>2</sup>.

As defined in [7], Semantic Proximity values were calculated for all pairs of articles and associated keywords, resulting in a SKS graph for each user cluster. Fig. 1 shows a graph that represents the keywords relational information for the community of users in cluster 3, i.e a measurement of this group's SKS.

After having derived a set of different SKS like the one shown in fig. 1, we can derive a set of general properties for SKSs (e.g. metricity, transitive closure, etc.) and consequently automate their derivation from digital library web resources. From this preliminary data, we plan to generate a set of agents according to the methodology described in section 1.3, and use these for a first series of simulations.

## 3 Conclusion

We presented a methodology that based on the concept of shared knowledge structures will enable us to measure and describe the performance of active recommendation systems.

Future research will focus on applying this methodology to a number of existing active recommendation systems that can lead to the following applications. First, this evaluation methodology can be used to gather empirical data on the knowledge of communities of researchers within an institution. Second, measurements of the speed of convergence, the suitability for specifically structured SKS, etc. will allow researchers to

<sup>2</sup> this process will be further detailed at the ECDL2000 conference presentation

