

User Requirements Report (D1.1)

Deliverable Type: R
Number: D.1.1
Nature: Report
Contractual Date of Delivery: month 3
Actual Date of Delivery: 28/05/2001
Task: WP1

Name of responsible: Dimitris Plexousakis

Editor : Dimitris Plexousakis
Institute of Computer Science
P.O. Box 1385
Vassilika Vouton
71110, Heraklion
Greece
dp@ics.forth.gr

Quality Insurance table

Revision	Modification Reasons	Date	Modifications Author's Visa	Deliverable Responsible's Visa	Project Manager's Visa
E	Final Validation	29/05/01	Ronchaud	Plexousakis	Straccia
D	Minor corrections	28/05/01	Ronchaud	Plexousakis	Straccia
C	Adjustments	28/05/01	Plexousakis	Plexousakis	Straccia
B	Presentation Modifications	23/05/01	Ronchaud	Plexousakis	Straccia
A	Creation of Document	23/05/01	Plexousakis	Plexousakis	Straccia

Abstract: The User Requirement document is the first tool on which to rely in order to determine the Cyclades system specifications. In order to gather the User Requirements, efforts have been put first on identifying the user communities. Once clearly identified, the second phase consisted in assessing their expectations as users of the Cyclades environment.

Keyword List:

User Requirements, identify user communities, expected functionalities, system design

Table of Contents

<i>Defining User Requirements</i> _____	3
Objectives of Workpackage 1 _____	3
Identifying Scholarly User Communities _____	3
Questionnaire Preparation _____	3
<i>Analysis of Questionnaire Results</i> _____	4
Structure of the Questionnaire _____	4
Questionnaire Results _____	4
Respondent's Profile _____	4
Questionnaire Feedback and Recommendations _____	5
Appendix A: User Requirements Questionnaire _____	7
Appendix B: Collected Results and Statistics _____	13

Defining User Requirements

Objectives of Workpackage 1

The objective of Workpackage 1 is to identify relevant user communities, select a user requirements collection and analysis methodology and collect and analyze user requirements regarding the CYCLADES environment. The resulting requirements will be used in Workpackage 2 for devising detailed functional specifications of the system while taking into account user needs.

Identifying Scholarly User Communities

A number of scholarly user communities were identified as “relevant” for providing feedback regarding the envisioned environment. These communities include:

- the German Society of Physics
- the German Society of Mathematics
- the Italian Society of Mathematics
- the Delos Network of Excellence in Digital Libraries
- the PLANET Network of Excellence in Artificial Intelligence Planning
- the Open Archives Initiative group

Academic and research groups among the partners’ institutions have also been consulted.

Questionnaire Preparation

The next step consisted in creating the specific questionnaire for eliciting requirements from prospective users of a collaborative archive environment. All partners contributed to the formation of the questionnaire by specifying appropriate questions for the different functional modules of the CYCLADES environment.

Considering the different disciplines of the user groups we intended to address with our questionnaire, an abstraction of technical details was made in order to produce a set of questions that sufficiently covered the entire system functionality, but which were presented in understandable terms by the scientific community at large.

The questionnaire also included a concise description of the CYCLADES objectives and of its different service components. Several versions of the questionnaire were produced, refined as well as modified by the CYCLADES partners. It should be noted that the questionnaire was also disseminated to the Open Archives Initiative group, whose valuable feedback was accommodated in the final version of the questionnaire.

The final version of the questionnaire is included in Appendix A.

It was decided to electronically disseminate the questionnaire in order to avoid delays with mailing printed versions. The questionnaire was made accessible via the World-Wide Web on a site hosted by ICS-FORTH.

(URL: http://sappho.ics.forth.gr:5000/User_requirements/questionnaire.html).

Solicitation letters were sent by the partners to the above mentioned communities. A total of 34 answers to the questionnaire were collected through the web site. An analysis of the results is presented in the following section and the complete set of answers and related statistics is included in Appendix B.

Analysis of Questionnaire Results

Structure of the Questionnaire

The questionnaire was structured as follows:

- introduction to the CYCLADES environment and its objectives
- user information section (6 questions)
- CYCLADES services sections
 - Search & Browse Service (8 questions)
 - Personalization Service (4 questions)
 - Recommendation Service (7 questions)
 - Collection Service (4 questions)
 - Collaborative Work Service (5 questions)

The questionnaire comprises a total of 34 questions distributed as evenly as possible among the different services. A particular effort was made to keep the questionnaire as succinct as possible.

Questionnaire Results

Respondent's Profile

The majority of the results obtained were collected from academics and only a small percentage from the industry. More than half of the respondents already belong to some Web Community or Scientific Network and have experience in using collaborative work support tools. All of them also use digital document archives for their work/research needs regularly (62%) or occasionally (38%).

Questionnaire Feedback and Recommendations

The majority of respondents find the functionality provided by the digital document collections they have used insufficient. Furthermore, only 27% of the people consulted appear to be willing to tolerate having to use specific client-side software, whereas the majority (73%) prefer the use of an unmodified browser.

Regarding the Search & Browse Service, half the respondents do not wish to use schemas other than Dublin Core in searching and browsing, whereas proposal for other metadata schemas include MARC and XML Schema. The vast majority of answers are expressing an interest for using attribute value – based search and there is a significant interest in specifically searching for person names, dates as well as free text. Half of the respondents wish to query the system using predefined fields; fewer prefer using a single field and only a small percentage (9%) appear to be willing to use a formal query language. A strong preference for the ability to formulate Boolean combinations of query conditions is expressed.

As far as the ordering of query results is concerned, relevance and publication date are the preferred modes. Half the respondents would occasionally use a query result rating feature for the purpose of improving search. The vast majority of respondents wish to have the ability to browse attribute values appearing in query result sets or restrict results further by additional queries.

In the Personalization section of the questionnaire, 44% of the respondents wish to use the CYCLADES environment as single scholars, whereas 41% wish to use it as both single scholars and as members of user communities or groups. Only 15% wish to use it only as part of their community or group activity. The majority of prospective users wish to subscribe their interests to the system so as to be notified of relevant additions. Furthermore, the majority of respondents wish to receive recommendations by the system as far as the organization of their thematic folders is concerned, but do not necessarily want this reorganization to take place automatically: 42% wish to perform this manually.

Recommendations of both documents and users / communities with similar interests are deemed useful by the questionnaire respondents. An almost identical percentage would be willing to rate documents so that other users can benefit from them. Users by far prefer a wide range of rating values over a narrow one. In addition, prospective users wish the system to interpret their actions as signs of interest or non-interest in a document, whereas only half of them approximately wish to have their identity given away by the system as the originators of a document rating. The majority would accept being recommended to other users as “users with similar interests”. All the respondents wish to be informed of a community’s topics when they receive community recommendation, where as approximately half of the respondents additionally want to see the most popular documents of the community and the name of the community.

In the Collection Service part of the questionnaire, respondents wish to define collections either by refining existing collections or by composing them. A very small percentage prefers to use predefined collections only. Half of the respondents favor a list of simple conditions on attribute values as the collection description. Fewer favor a textual description and a small percentage suggest the use of graphics and icons. Questionnaire fillers were also asked to produce descriptive sentences of their collections of interest. The responses ranged from very specific collection descriptions (e.g., “Epilepsy, EEG, Nerve Conduction studies”) to very general ones (e.g., “computer science, library and information science”). Finally, the majority of respondents prefer searching and browsing on specific metadata fields or with the use of specific terms, rather than performing searching and browsing with returned records in a special format.

The last section of the questionnaire covered the Collaborative Service. 46% of the respondents wish to see annotations only on request, whereas a slightly smaller percentage wish to always see the subject line of annotations with the document. The majority of respondents want to receive notifications about changes to shared documents according to their preferences and also wish to receive name and e-mail information about other users. They are willing to share the same type of information with other users of the system. The availability of a chat tool to communicate with other users also appears to be desirable, although do not deem the provision of such a tool important.

Appendix A: User Requirements Questionnaire

Project CYCLADES: User Requirements Questionnaire

By The CYCLADES Consortium

April 2001

Foreword

This questionnaire has been devised for the elicitation of user requirements for the project CYCLADES (IST-2000-25456). A brief description of the project and its objectives is given below. The CYCLADES consortium aims to elicit potential users' views on the functionality and services provided by the CYCLADES environment. Your effort in filling out this questionnaire is greatly appreciated.

Project Description

The main objective of CYCLADES (<http://www.iei.pi.cnr.it/cyclades/>) is to develop *advanced Internet accessible services to support scholars both individually and as members of networked communities when interacting with large interdisciplinary electronic (e-print) archives*. CYCLADES aims at supporting the transition of e-print systems into genuine building blocks of a transformed scholarly communication model by developing a set of leading edge technologies providing innovative methods for information access, dissemination, sharing and collaborative work.

The proposed open archives environment consists of two components: the archives and the services. The former will participate using an interoperability protocol developed by the Open Archives initiative (OAI). This protocol enables archives to expose metadata in various forms and that can be used by a variety of services. CYCLADES will base the development of the service environment on these specifications. In particular, a core set of cross-archive value-added services will be developed to constitute a federation of independent but interoperable services. The Service Environment will provide OAI compliant functionality. The CYCLADES services comprise the following:

- *Access Service*: supports information gathering, plus indexing and storage of gathered information in a local database.
- *Search and Browse Service*: develops plans for the execution of user queries. An ad-hoc or a profile-based user query will be decomposed into more simple sub-queries to be sent to the Access Service for execution. The results of the sub-queries are fused and returned to the user. A browse facility is also supported.
- *Collection Service*: provides mechanisms for dynamically building meaningful collections.
- *Personalization Service*: supports personalization of information access on the basis of individual user profiles and of profiles of scholarly communities to which users belong.
- *Recommendation Service*: provides recommendations to satisfy information needs of a user based on ratings provided by other users or groups.
- *Collaborative Work Service*: supports collaboration between members of virtual communities. Community working areas are created to use the OAI content in collaborative work.

User Info

- 1 What is your area of work? Academia Industry Research
- 2 Are you a member of any Web Community or Scientific Network? Yes No
- 3 Have you ever used a collaborative work support tool (e.g., Lotus Notes, Hyperwave, BSCW) ? Yes No
- 4 Do you use digital document archives for your work/research needs ?
Never Occasionally Regularly
- 5 Do you find the functionality provided by the digital document archives you have used sufficient?
Absolutely not Could be improved Satisfactory in most cases
- 6 Would you prefer accessing the Cyclades environment using an unmodified Web browser, or could you also tolerate the necessity of client-side software installation?
Prefer unmodified browser Would also use specific client-side software

The CYCLADES Open Collaborative Archive Environment – Services

The questions that follow refer to the specific services that the CYCLADES environment will deliver. Each service is briefly explained in the respective questionnaire section.

Search and Browse Service

This service provides the search and browse facilities for a set of archives compliant with the Open Archives Initiative Specifications. We distinguish three levels of abstraction for which searching and browsing will be supported:

- The metadata level contains the metadata records of documents. Searching allows for querying for metadata records fulfilling a combination of query conditions, where each condition refers to the value of a specific attribute in a metadata record (e.g. creator='Smith', title='searching open archives').
- The next higher level is the attribute value level. A user might want to look at possible attribute values without retrieving and reading the whole records. For example, a user might want to know, if there is an author named "Fuhr" at all, or if there are authors whose names sound like "Fuhr".
- The highest level of Multilevel Hypertext is the schema level. The Open Archives specification requires Dublin Core as a standard metadata schema, but it allows also other schemas to be provided additionally. When formulating a query, a user might want to use a more specific schema than Dublin Core, and then, she would have to know which schemas are available and which attributes (fields, properties) they provide.

- 1 Apart from Dublin Core, do you want to use other schemas in searching and browsing?
I don't care about metadata schemas / I don't use metadata schemas
No
Yes (please name schemas):

2 Do you want to search for attribute values? (E.g.: Would you like to be able to look for author names, publication dates, publishers etc., independently of individual documents?)
Yes No

3 What types of data do you need to distinguish when formulating a query? (E.g. would you search explicitly for persons, or would it be sufficient to search for words in general to find a person?)
String Free text Person names Dates Numeric values
Other (please specify) : I don't care about data types

4 How would you like to query the system?
Fill in predefined field boxes (as in most digital libraries)
Use a formal query language (e.g. like SQL)
Enter search values in a single field (as e.g. in Web search engines)

5 When you use multiple conditions in a query (e.g. *date*=2000 and *author*=Smith), how should these conditions be combined?

A list of optional conditions only (e.g. with the conditions *date*=2000 and *author*=Smith, you would get result where the date is 2000, or where the author is Smith)

A list of conditions, each marked as optional, mandatory or negated (e.g. you could ask for documents that have the author Smith, maybe the date 2000, and were not published by UniDo, by the following list: mandatory:*author*=Smith, optional:*date*=2000, not:*publisher*=UniDo)

Boolean combination of conditions (e.g.: *author*=Smith AND *date*=2000 AND NOT *publisher*=UniDo)

I don't understand the question

6 For query results, which kinds of ordering should be supported:
Relevance (I.e.: How well do the result items match the query? Show the best matches first.)
Publication date
Other (please specify):

7 Some systems let you rate the documents retrieved by a query and use this rating (i.e., your opinion about the usefulness of the document, e.g. "useful" versus "not useful") to improve the query automatically. Would you use such a feature?
Yes, probably often
Yes, occasionally
No, never
I am not sure, I have never used rating resp. relevance feedback before

8 Which operations should be supported for query results:
Browse attribute values occurring in result set
Restrict result set by another query
Other (please specify):

Personalization Service

This service is responsible for supporting personalised information provision based on automatically acquired profiles of users; users are viewed both as individuals and as members of communities sharing common interests. User profiles are organised hierarchically into topic folders, similarly as in common e-mail programs. New information will be automatically classified into the right user topic folder.

- 1 How do you view your usage of the system?
As a single user/scholar As a member of a Community / Group Both
- 2 Would you be willing to subscribe your interests to an evolving archive so as to be notified of possibly relevant additions? Yes No
- 3 Would you like the system to automatically subdivide search results into hierarchically organized thematic folders? Yes No I prefer to do this manually
- 4 Would you like the system to suggest how to reorganize your hierarchy of thematic folders?
Yes No

Recommendation Service

This service delivers recommendations about newly published documents that might be of interest to the user. The recommendations are based on the ratings which the documents received from other users. The system also recommends communities of other users that have similar interests and information needs.

- 1 How useful are the different kinds of recommendations?
recommendation of documents:
 very useful useful less useful
recommendation of users/communities with similar interests:
 very useful useful less useful
- 2 Would you rate documents so that other users with similar interests can profit from your ratings? Regularly Frequently Rarely Never
- 3 Recommendations are based on ratings of documents. What range of values would you prefer when rating?
Narrow (e.g. irrelevant - relevant)
Wide (e.g. irrelevant - don't know - interesting - very interesting - most interesting)
- 4 Would you tolerate the fact that the system tries to interpret your actions, e.g. deleting a document, reading it, printing it, as signs of interest or non-interest in a document?
Yes No
- 5 Would you allow the system to recommend you as a "user with similar interests" to other users or communities and pass on your name and possibly e-mail address? & Yes & & No

- 6 Would you allow the system to give away your identity to other users as the originator of a document rating? Yes No
- 7 What information would be suitable in describing a new community that is recommended to you by the system? (check one or more)
- The name of the community
 - The most popular documents of the community
 - The community's interest given as topics and keywords

Collection Service

The Cyclades system will provide the user with a view of an information space structured into a set of virtual collections. This structure will be dynamic, i.e. new collections can be added to satisfy the needs of the users. Users will be able to both restrict their information space to those existing collections that fit their information needs and to create their own collections. Each collection will be characterised by a name and a description that specifies the nature of the documents contained in the collection. The users will select their collections by relying on this description. Possible examples of collection descriptions are: "Documents in the domain of 'Mathematical models for forecasting' ", "Thesis and dissertations produced by students at X,Y and Z departments", "e-print documents in the domain of 'Computer Science applied to Economy' that are written in French".

- 1 How would you like to define your collections?
- As composition of existing collections
 - As refinement of existing collections
 - I would prefer to use only predefined collections
- 2 In which description format would you like to see the description of a collection when you select your information space?
- As a textual description
 - As a list of simple conditions "attribute = value" (e.g. subject= historical films, temporal coverage = 1936-1956)
 - Other (please specify)
- 3 Please, describe your collections of interest as short descriptive sentence

Collection specific search and browse services will be provided, when possible, in addition to search and browse on the common metadata set (Dublin Core). For example, a browse on the DIF metadata format might be provided for the collection "South Pole geographic documentation", a search on a subject selected from MESH might be associated with the collection "Journal Articles on Diabetes Pathologies", and a search that returns US Marc records might be returned as result of searches on "English literature of the seventies".

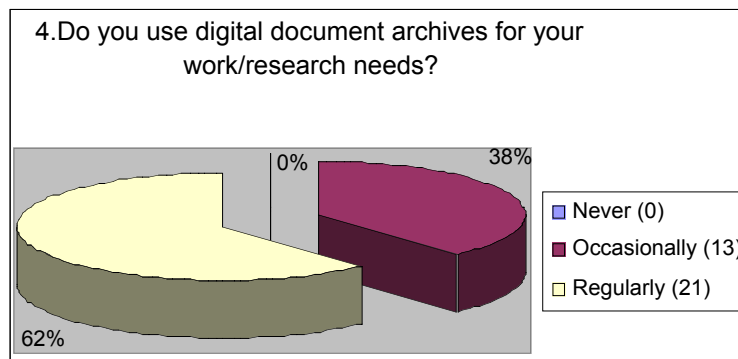
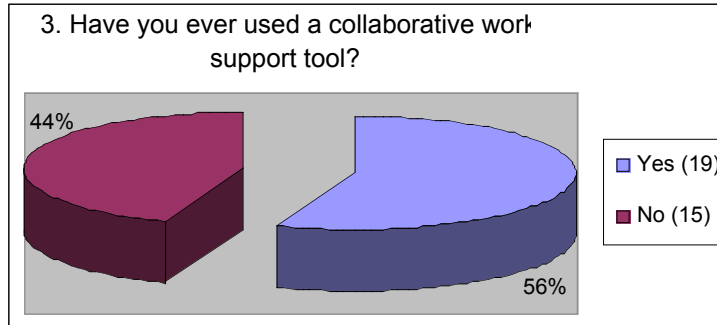
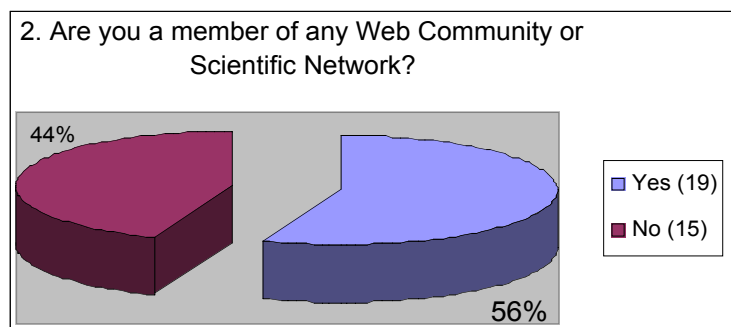
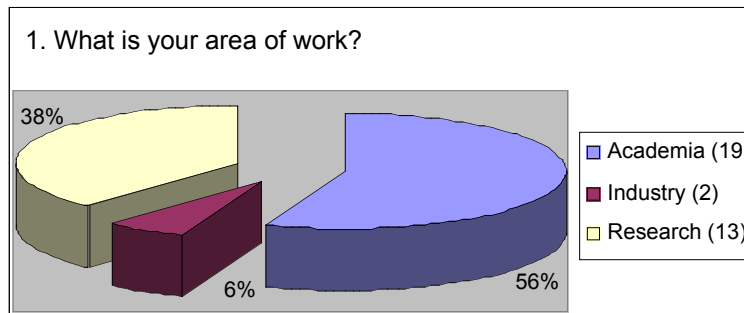
- 4 What collection specific search and browse do you think are useful?
- Search/browse on specific metadata fields
 - Search/browse with specific terms
 - Search/browse which returns records in a special format (e.g., Dublin Core, MARC, etc.)

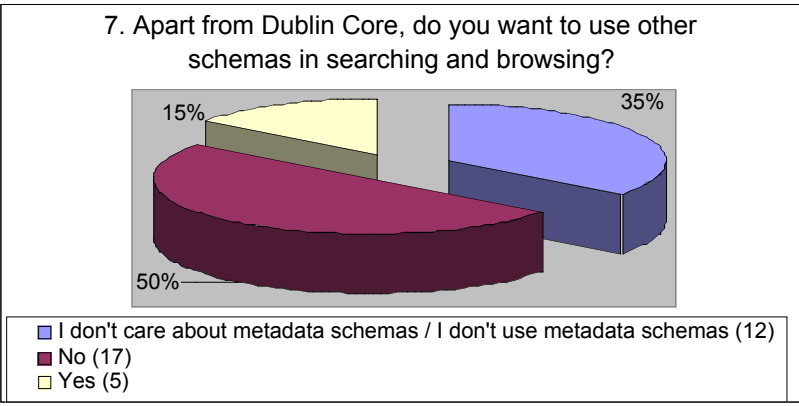
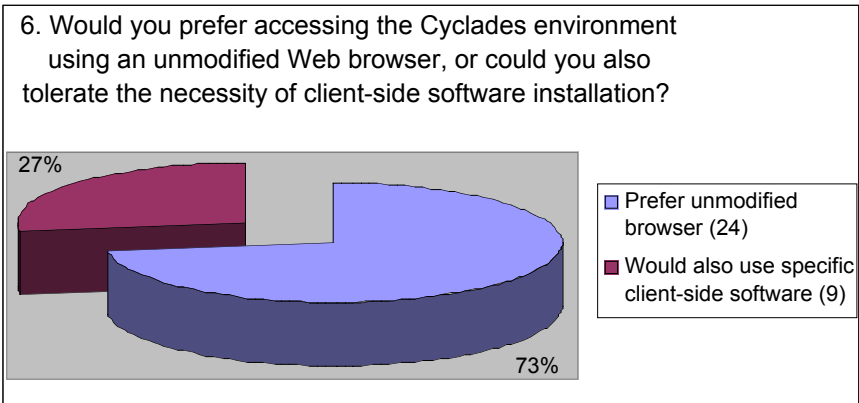
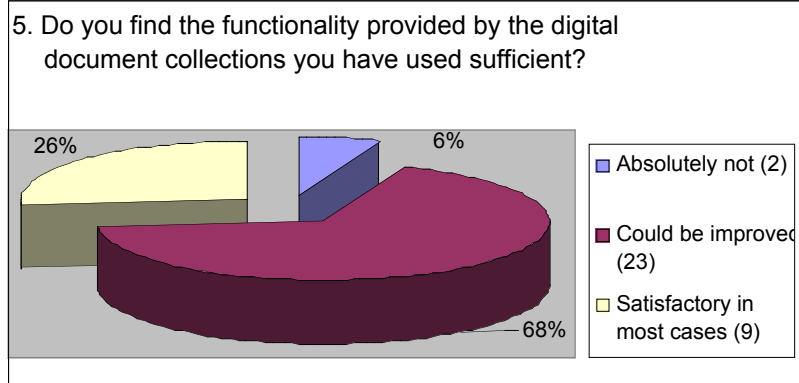
Collaborative Work Service

The Cyclades system allows users with similar interests to set up and maintain communities of interest. In these communities users can share documents, annotate and discuss documents, and receive notifications about changes. Furthermore, users get information about other users who are online and can contact them with a chat tool.

- 1 How do you want annotations to be displayed?
- See whole annotation always with document
 - See subject line of annotation always with document
 - See annotations only on explicit request
- 2 When would you like to receive notifications about changes to shared documents?
- Immediately Upon login to the system
 - Once a day According to my personal specification
- 3 What kind of information about other users would you like to receive (multiple selection possible)? Name Email address Snail mail address
 Telephone and fax number Other:
- 4 What kind of information would you be willing to share with other other users (multiple selection possible)? Name Email address Snail mail address
 Telephone and fax number Other:
- 5 Would you use an integrated chat tool to contact other users? Yes No

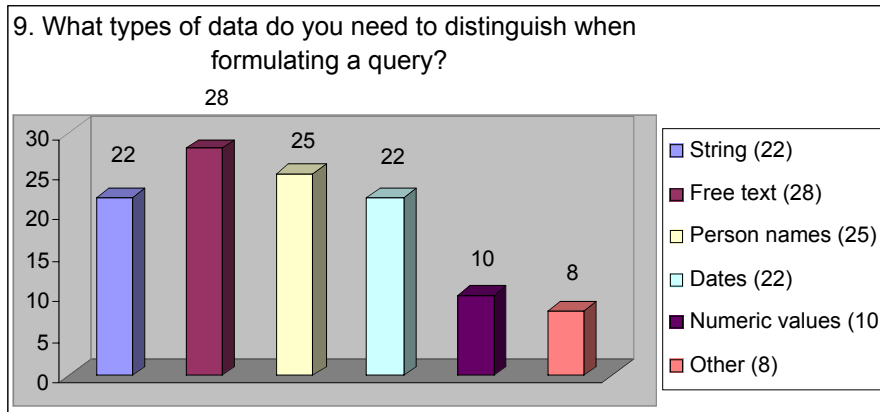
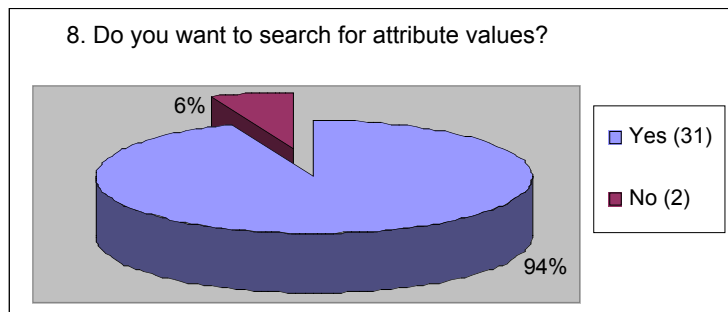
Appendix B: Collected Results and Statistics





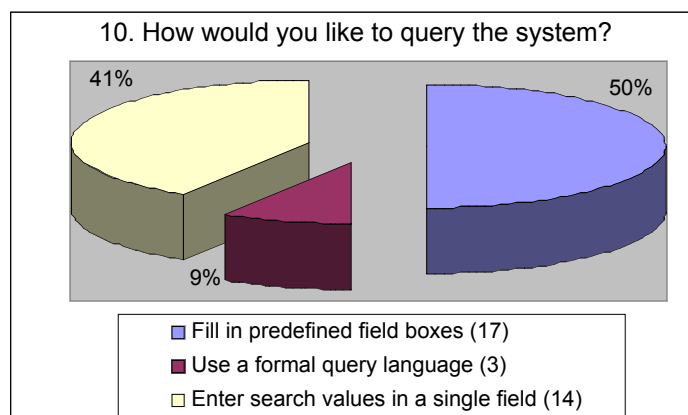
Please name schemas:

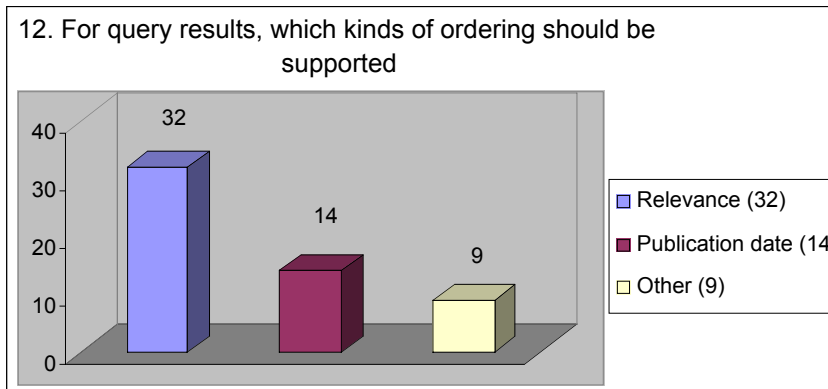
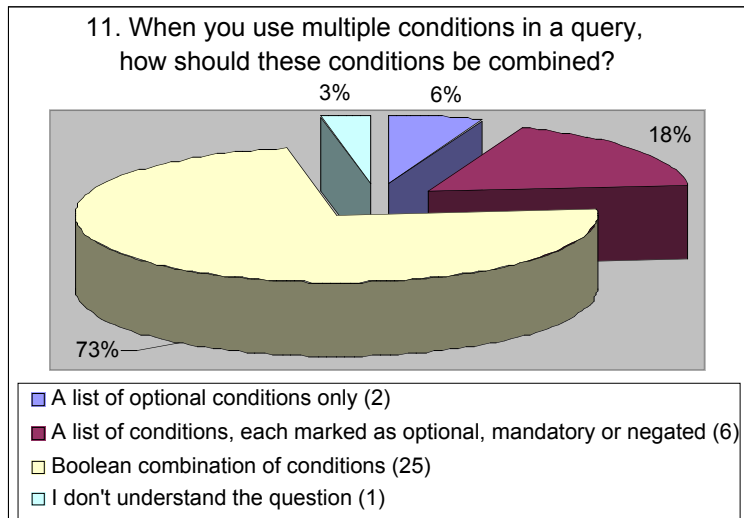
- MARC records
- MARC
- Don't know yet, maybe some of the upcoming XML Schemas etc.
- I don't know enough about schemas yet, but I might use them in the future
- Attribute value level



Other:

- Exact string
- Subject
- Keyword/ index
- Searching in subject and keywords
- Classification codes
- Boolean and position of string
- Text in different languages

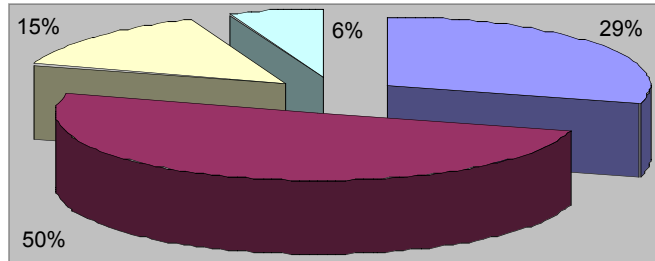




Other:

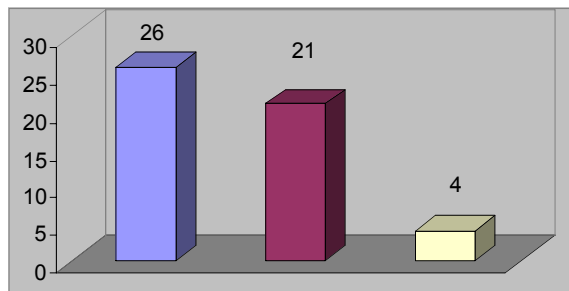
- Collection
- The meaning derived from the words together in the right context
- By subject
- (Read) Popularity / Ranking / Rating
- Sorted by creator name
- Subject/Category
- Author
- Peer reviewed or not
- Title

13. Some systems let you rate the documents retrieved by a query and use this rating to improve the query automatically. Would you use such a feature?



- Yes, probably often (10)
- Yes, occasionally (17)
- No, never (5)
- I am not sure, I have never used rating regarding relevance feedback before (2)

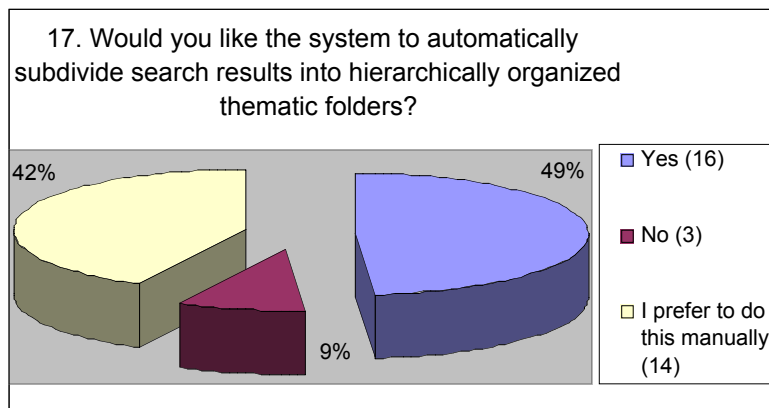
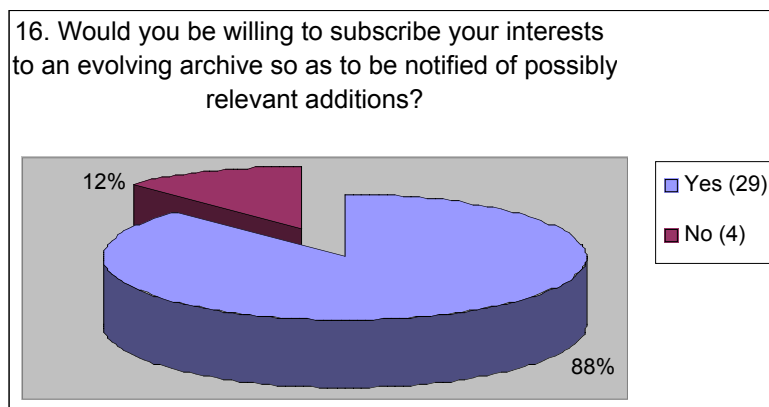
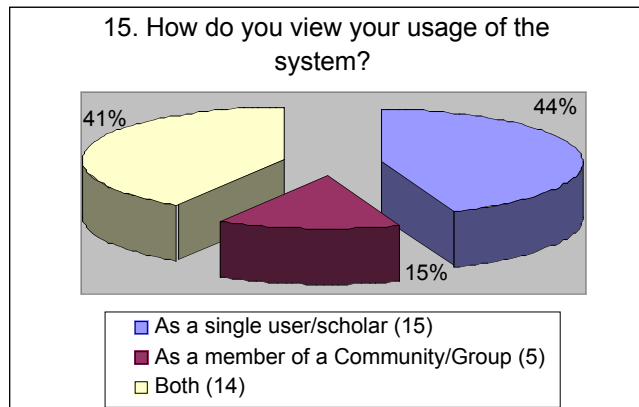
14. Which operations should be supported for query results?

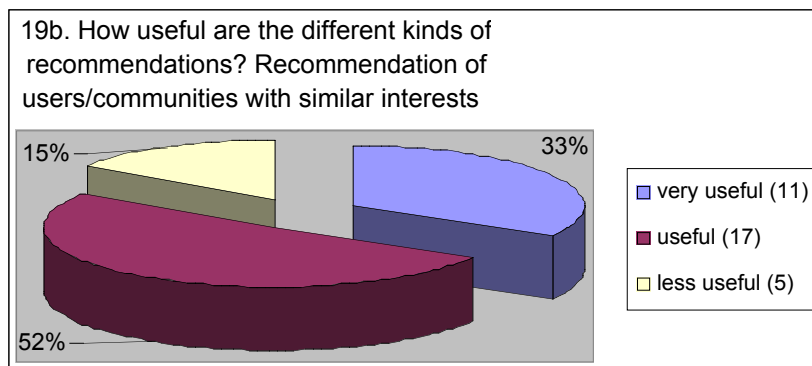
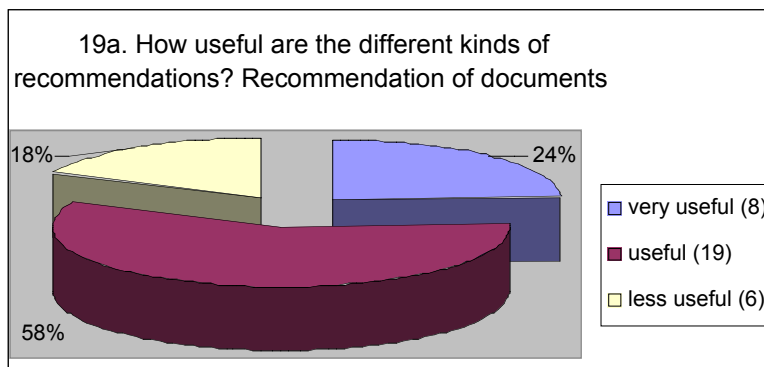
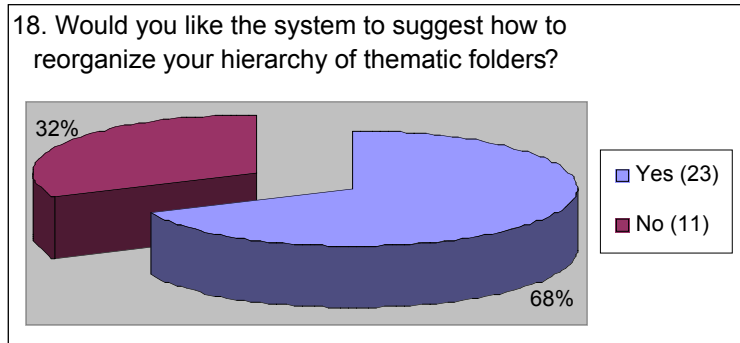


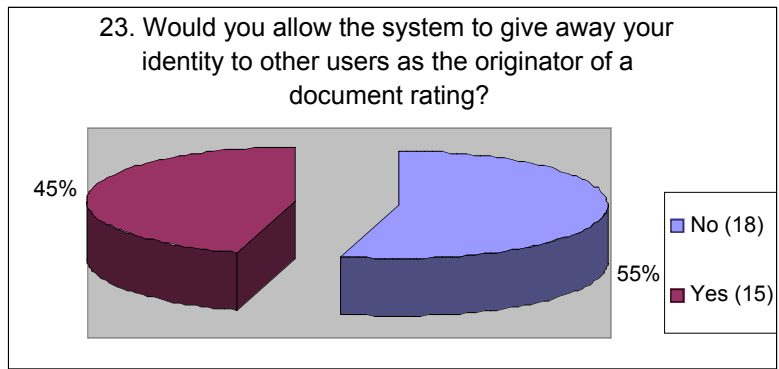
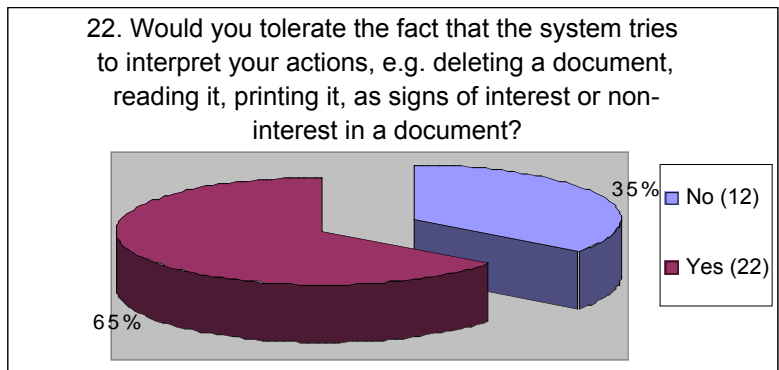
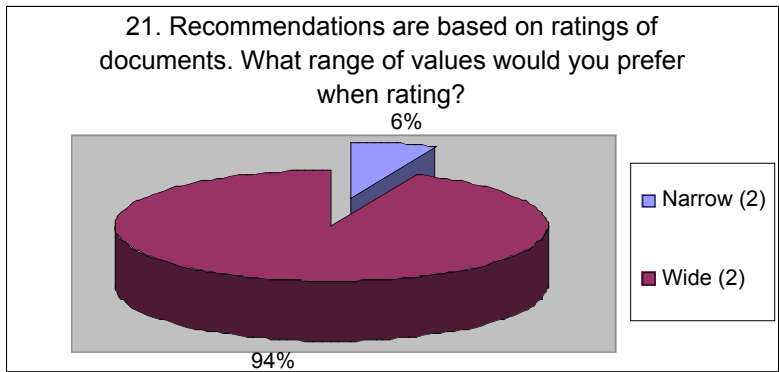
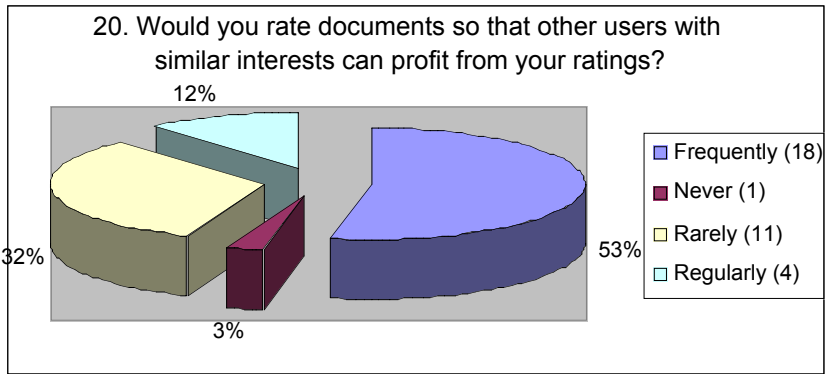
- Browse attribute values occurring in result set (26)
- Restrict result set by another query (21)
- Other (4)

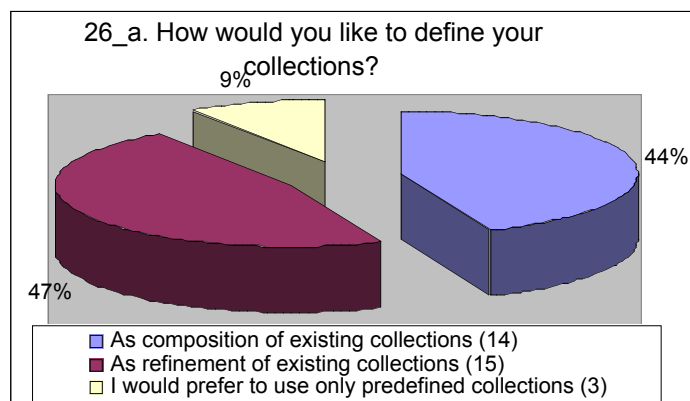
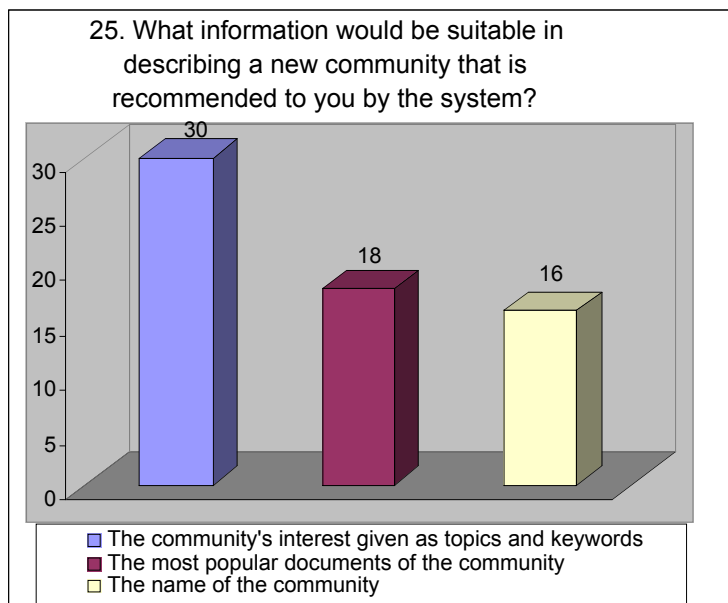
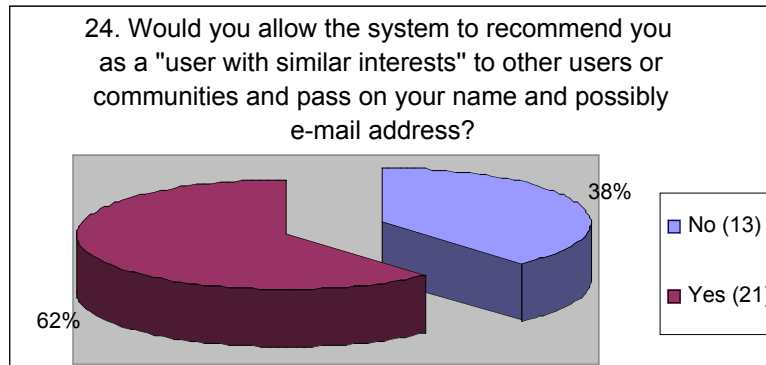
Other:

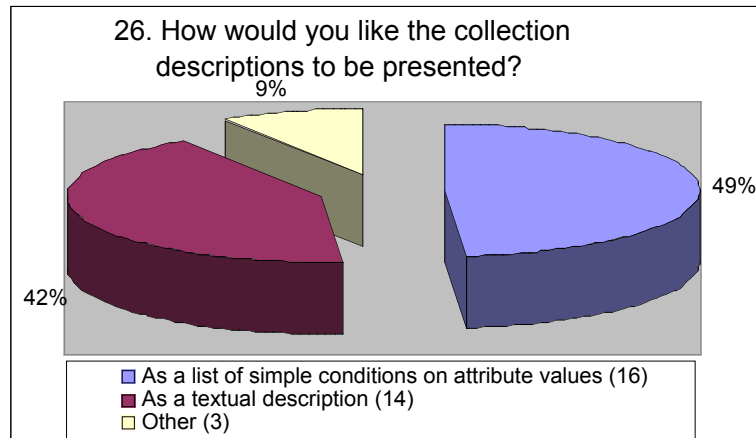
- Saving / book marking result sets
- Make more specific; make more general
- Find similar objects, relevance feedback queries
- Browse metadata, TITLE, for example









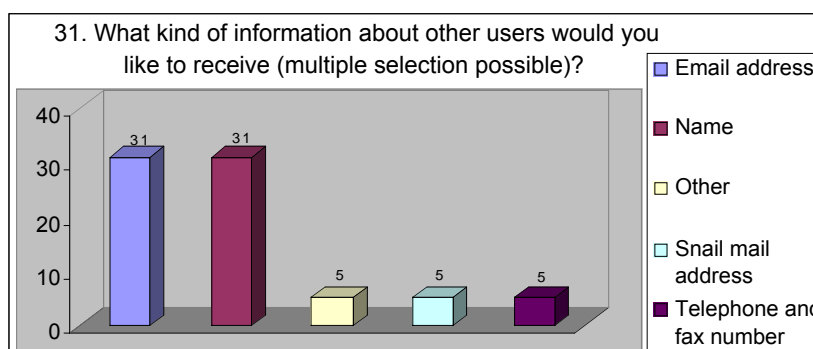
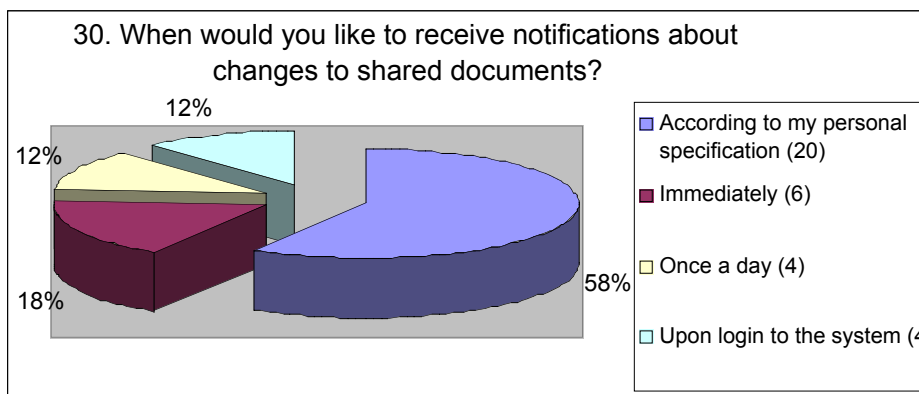
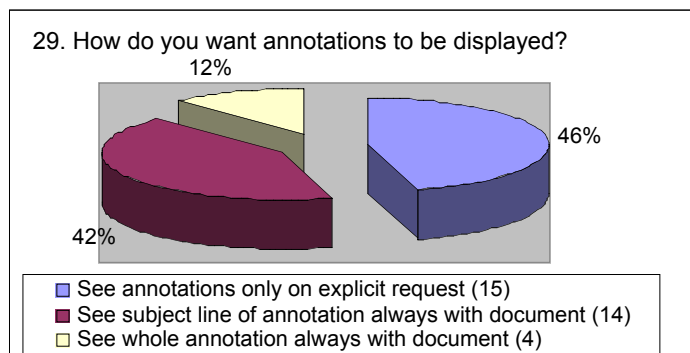
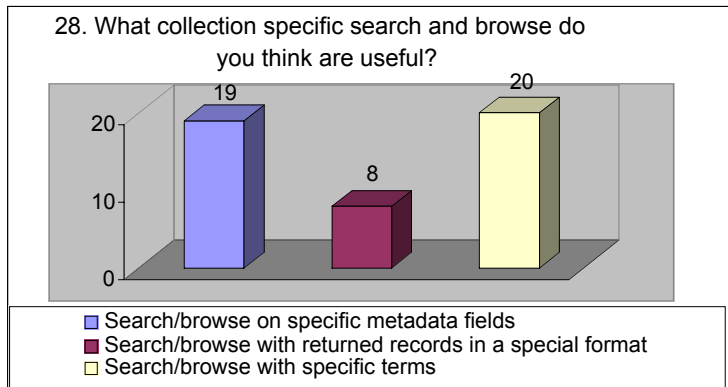


Other:

- Maybe a graphic visualization would be useful? Possibly a Venn diagram. Otherwise, I want to choose -- maybe today I feel like text, tomorrow like list of simple conditions
- Textual and maybe graphics / icons
- Both

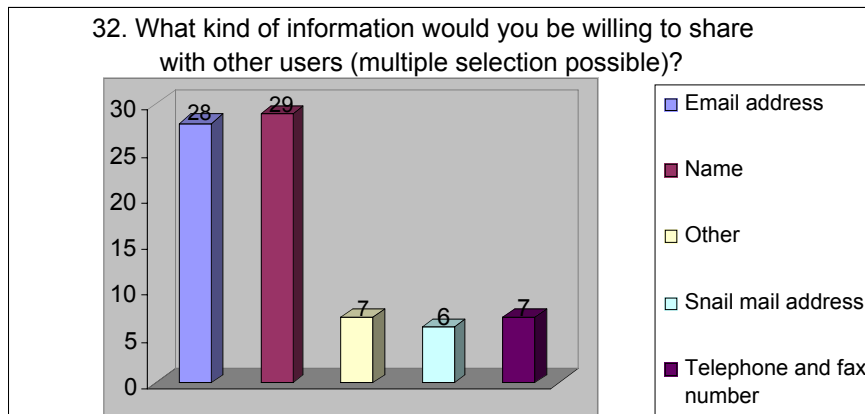
27. Please, describe your collections of interest as short descriptive sentence.

- Computer science, library and information science
- Scientific communication
- "Historical films on the Greeks"
- Journal articles and essays in philosophy, especially Peirce, Jaames, Dewey, Randall, Buchler, Ernest Nagel.
- Digital library development particularly issues related to users such as document creation and self-archiving.
- Integration of IR and DBMS.
- Books on CSCW, latest conferences and workshops in my field.
- CSCW and the web
- Large Scale Structure, Simulations, Hybrid
- Documents about brand equity measurement.
- Scholar communication
- Epilepsy, EEG (electroencephalography), Nerve Conduction Studies, EMG (electromyography).
- Information sciences, digital libraries



Other:

- Field of practice/research/occupation
- Which institution they come from; e.g. University of Copenhagen
- Communities, Personal Description / Profile, ICQ & IM, Homepage, Since when active
- Academic field
- Home page on the web



Other:

- Professional profile / url
- Field of practice/research/occupation
- My institution
- The system should provide a way to send an email without having to know the email address in advance
- Email on demand / protected) Communities, Personal Description / Profile, ICQ & IM, Homepage, Since when active
- Academic field
- Home page on the web

