
An Agent-based System for Modelling the Searching Process on the Web

Cândeia, C.

AI Research Group
Wittmann & Partner Comp. Sys.
2400 Sibiu, ROMANIA
ciprinac@airg.verena.ro

Staicu, M.

AI Research Group
Wittmann & Partner Comp. Sys.
2400 Sibiu, ROMANIA
mariuss@airg.verena.ro

Zamfirescu, C.B.

Depart. Of Computer Science
“Lucian Blaga” Univ. of Sibiu
2400 Sibiu, ROMANIA
zbc@acm.org

Overview

1. Motivation
2. Objectives
3. IR and agents characteristics
4. Current trends in agent-based IR Systems
5. Design issues
6. SEA Architecture
7. User modelling
8. Interface agent
9. Application
10. Resource profiles
11. Mediators
12. Experimental results
13. Conclusions and future extensions

1. Motivation

- ✗ users may have no idea *where to start* their search,
 - ✓ *where to find* what they really want,
 - ✓ *what services are available* for doing their job,
- ✗ users may not *be aware of every change* in the Web,
 - ✓ services may appear and disappear over time
- ➔ user is simply overtaxed by manually searching the Web

2. Objectives

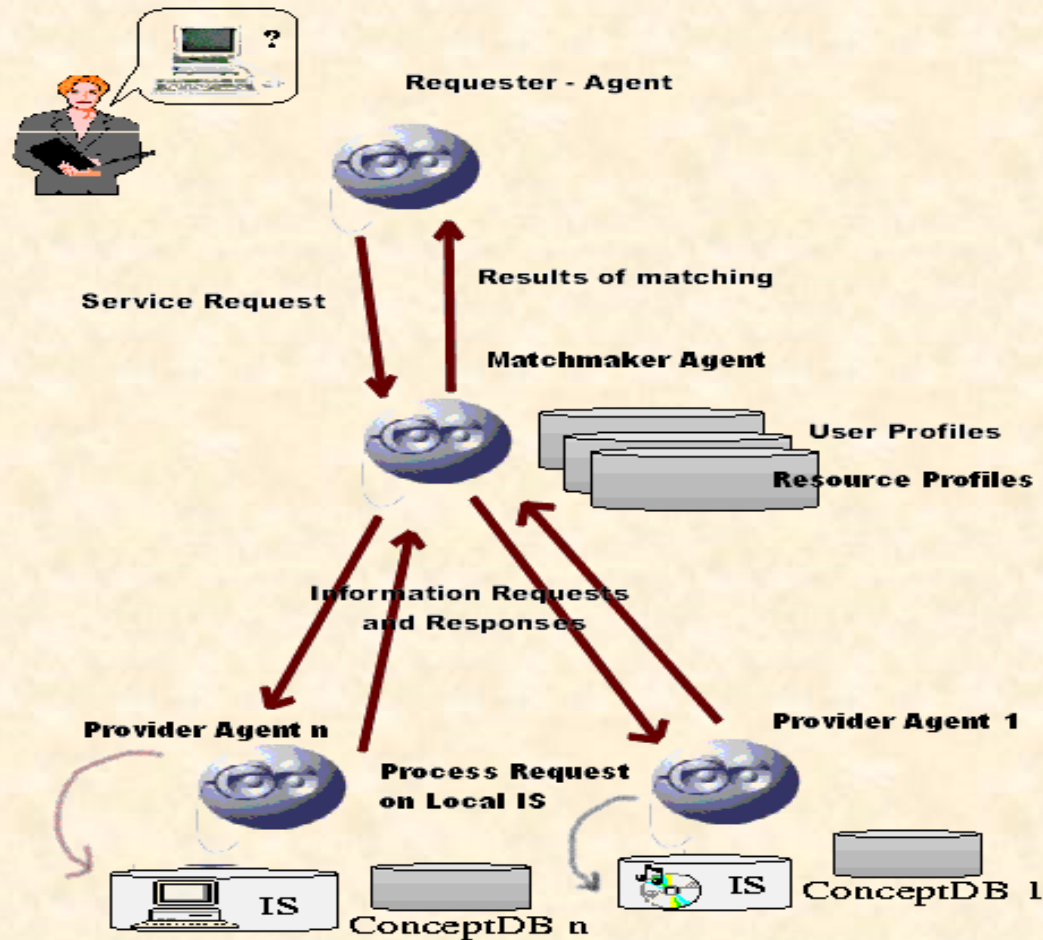
- ✓ assist the user in the diagnosis process and query reformulation
- ✓ select appropriate information source for an efficient searching
- ✓ translate the query accordingly to the available search engines
- ✓ manage searching strategy
- ✓ support the user in the results assessment
- ✓ provide the user with the appropriate outputs in a suitable form
- ✓ advice he/she in the follow-up activity

3. IR and Agent Characteristics

(Finin, Nicholas and Mayfield, 1998)

	A u t o n o m y	C o o p e r a t i o n	A d a p t a t i o n
Relevance Feedback		✧	✧
Information Extraction		✧	
Multimedia Retrieval		✧	
Effective Retrieval		✧	✧
Routing & Filtering	✧	✧	✧
Interfaces & Browsing		✧	✧
Term Expansion			
Efficiency & Flexibility	✧	✧	✧
Distributed IR	✧	✧	
Integrated solutions	✧	✧	✧

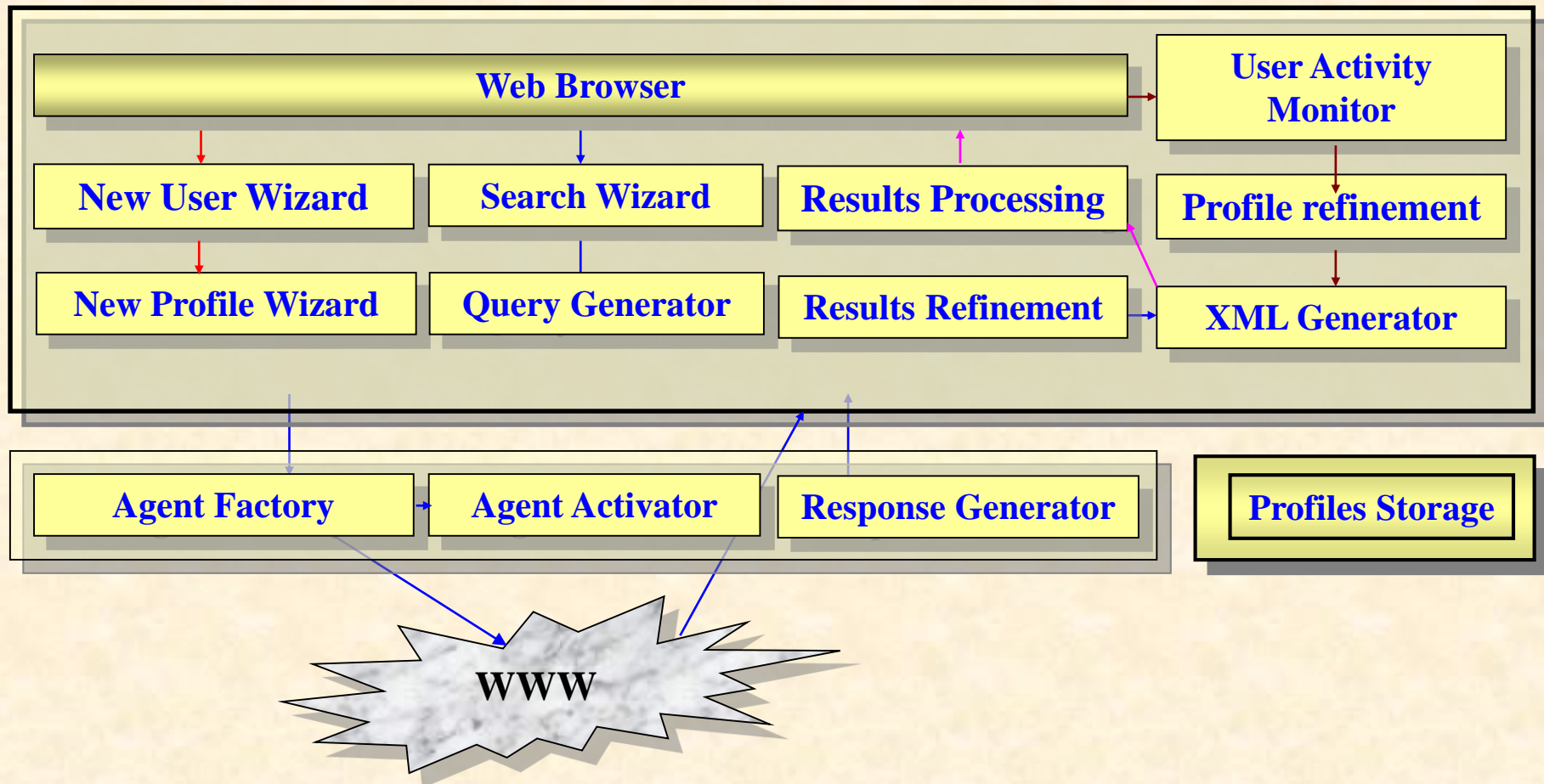
4. Current Trends



5. Design Issues

- ✗ To design a *flexible* and *extensible* architecture able to be easily adapted to any kind of:
 - ✓ *user's need*
 - ✓ *context of use*
 - ✓ *application domain*
 - ✓ *resource available*
 - ✓ *service provider*
 - ✓ *content*
- ✗ To be easily embedded in any kind of application

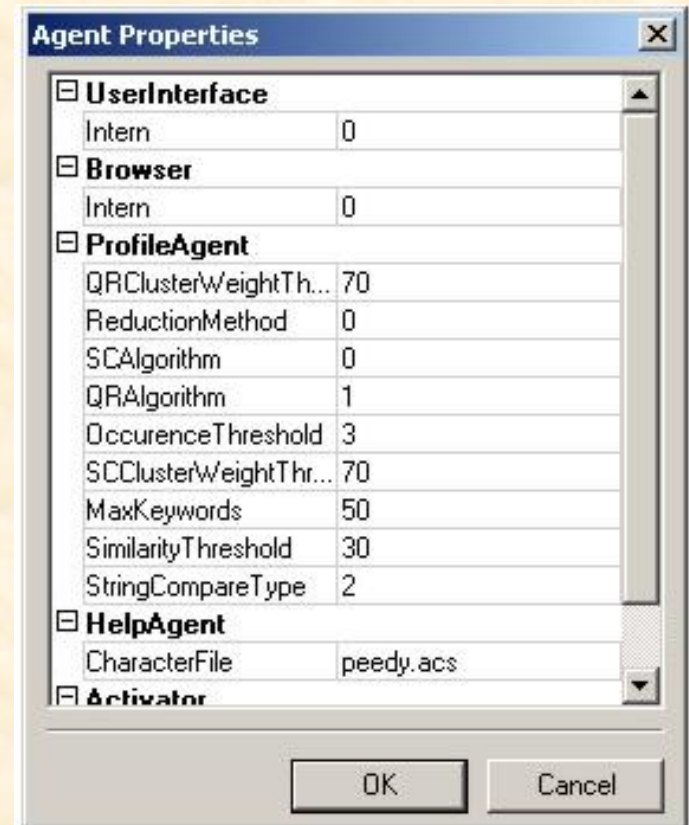
6. SEA Architecture



7. User Modeling

$$\left\{ \begin{array}{l} \text{Sim}(d_i, d_j) = 2 \times |d_i \cap d_j| / (|d_i| + |d_j|) \\ w_i = (0.5 + 0.5 \text{tf}(i) / \text{tf}_{\max}) (\log(n / \text{df}(i))) \\ p(w, m) = q(w) * m \\ u(w, m, s) = m + z(t) w \end{array} \right.$$

- w - a Web page
- m - user's interests
- $\text{tf}(i)$ - the term frequency
- $\text{df}(i)$ - the document frequency
- n - the number of documents
- tf_{\max} - the maximum term frequency
- $p(w, m)$ - the pertinence of a Web page to user's interest
- $q(w)$ - return the similarity measure
- $u(w, m, s)$ - updated user profile given the user's feedback s
- $z(t)$ - user's score for a page w



8. Interface Agent

New User Wizard

New Profile Information
On this page you will select the domains for the new profile

Name: Knowledge level: High

Description:

Domains:

- Information Technology
- Institutes
- Journals
- Libraries
- Life Sciences**
- Mathematics
- Measurements and Units
- Meteorology
- Museums and Exhibits

Selected Domains :
Research
Science and Technology Policy

New User Wizard

You have completed this wizard
If you press Finish the following changes will be made:

New User Results

- User
 - Name: nanu
 - Login: zbc
 - Email: zbc@aig.veneta.ro
- Profile
 - Name: gdss
 - Domains
 - Research
 - Science and Technology Policy

Search...

Request:

Search Quality :

Max. results: links

Results/search engine:

Search Timeout: sec.

Use history ☐

Use advanced setting ☐

Search Results

Search engine details:

	Search Engine	Timeout	Results
<input checked="" type="checkbox"/>	Excite	100	10
<input checked="" type="checkbox"/>	Google	100	10
<input checked="" type="checkbox"/>	InFind	100	10
<input checked="" type="checkbox"/>	Lycos	100	10
<input checked="" type="checkbox"/>	Northern Light	100	10
<input checked="" type="checkbox"/>	Yahoo!	100	10

User login

Login :

Profiles :

Name	Description
Comics and Animation	Comics and Animation
Genetic Algorithm	Genetic Algorithm

9. The Application

Sea Results - Sea2

File Edit View Favorites

Address: H:\Projects\SEA\Program\Debug\Response\responses.html

Search Engine Agents: airg.verena.ro

Search Results

Ratings

95%

mbox: **genetic algorithms + data structures = evolution programming** 99% - Directories Lists: mbox: **genetic algorithms + data structures = evolution programming** 99% - Directories Lists: mbox: **genetic algorithms + data structures = evolution programming** Toni Alatalo(antont@c Thu, 11 Sep 1997 18:36:48+0200)
Search Engine: Northern Light
<http://an.org/evolutio/0000.html>

59%

Genetic Algorithms in Stratigraphic Modeling 99% - Articles General info: Automation of stratigraphic simulations: quasi-backward modeling using **genetic algorithms**.(1) and(2) Geological Society Special Publication No. 134, 1998, eds: A. Mascle, C. Puigdefabregas,
Search Engine: Northern Light
<http://sun1.rzn-user.uni-hannover.de/nhdbkoqu/StratMod-GA.html>

59%

RC5 General Discussion: RE: [rc5] Genetic Algorithms 99% - Directories Lists: RE: [rc5] **Genetic Algorithms**. Chris Arguin(Chris.Arg Wed, 5 Nov 1997 21:35:56-0500(ES Messages sorted by: Next message: Previous

ProfileView DetailView

Sea Results

No	S.	Link	Description
1	95	http://an.org/evolutio/0000.html	mbox: genetic algorithms + data structures = evolution programming 99% - Director...
2	59	http://sun1.rzn-user.uni-hannover.de...	Genetic Algorithms in Stratigraphic Modeling 99% - Articles & General info: Automa...
3	59	http://lists.distributed.net/hypermail/r...	RC5 General Discussion: RE: [rc5] Genetic Algorithms 99% - Directories & Lists: R...
4	59	http://fuzzy.fzk.de/~rainer/new/work...	Automatic feature selection using genetic algorithms 99% - Directories & Lists: Rai...
5	47	http://seulab.phonon.ac.kr/1997/2001...	IMSL64C...A5.0.17A trajectory planner for manipulators using genetic 177%

Search History

Please, press 'This is it' when you find the info

Start Inbo... Wind... Sea ... Users... My D... Corel... IMP ... Sea ... EN 10:35 AM

Search Status

Status: Elapsed: 00:47

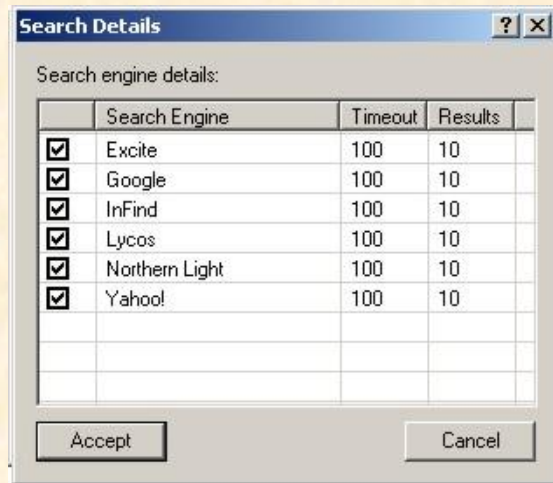
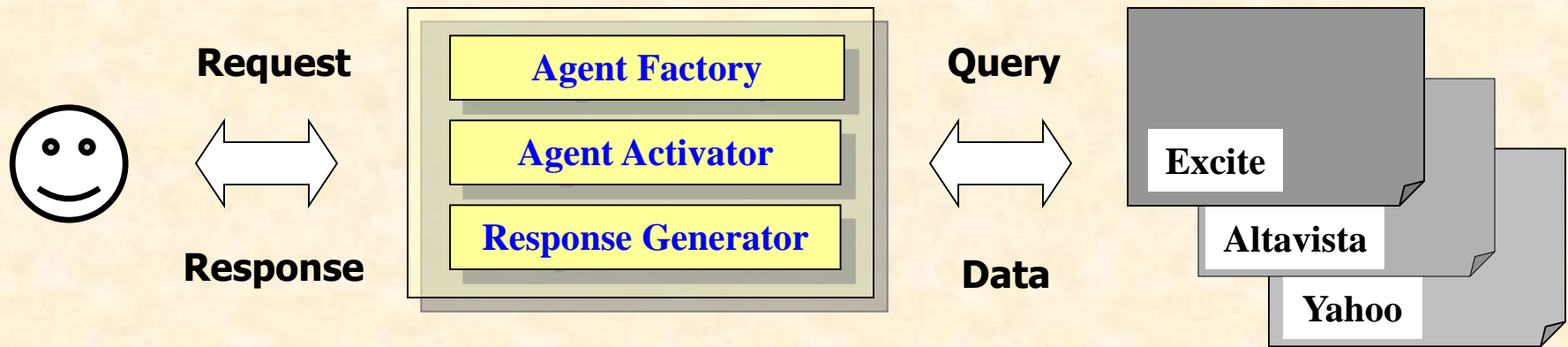
Hide Stop Details <<

Search Engine	Status	Timeout
Excite		253
Google		0
InFind		0
Lycos		0
Northern Light		0
Yahoo!		0

10. Search Engine Profiles

```
<sea type="search engine definition">
<description> <search_engine name="Yahoo!" url="http://www.yahoo.com" />
  <options> <timeout value="300" /> <results value="20" /> </options>
  <operators>
    <and value="+" /> <not value="-" /> <or value="" /> <exact_phrase value="yes" />
    <title_only value="t:" /> <url_only value="u:" /> <wildcard value="*" />
    <priority value="+-t:u:*" />
  </operators> </description>
<directories>
<directory name="The Web" code="00000000">
<params> <action value="http://search.yahoo.com/bin/search" inherit="yes" />
  <text name="p" inherit="yes" />
  <static name="h" value="s" inherit="yes" />
</params> <children> <directory name="Arts and Humanities" code="00000001">
<params> <action value="http://search.yahoo.com/bin/search" inherit="yes" />
  <static name="h" value="s" inherit="yes" />
  <static name="f" value="0%3A2766678%3A2718086%3A1" inherit="yes" />
  <static name="r" value="Arts" inherit="yes" />
  <static name="y" value="n" inherit="yes" /> </params>
```

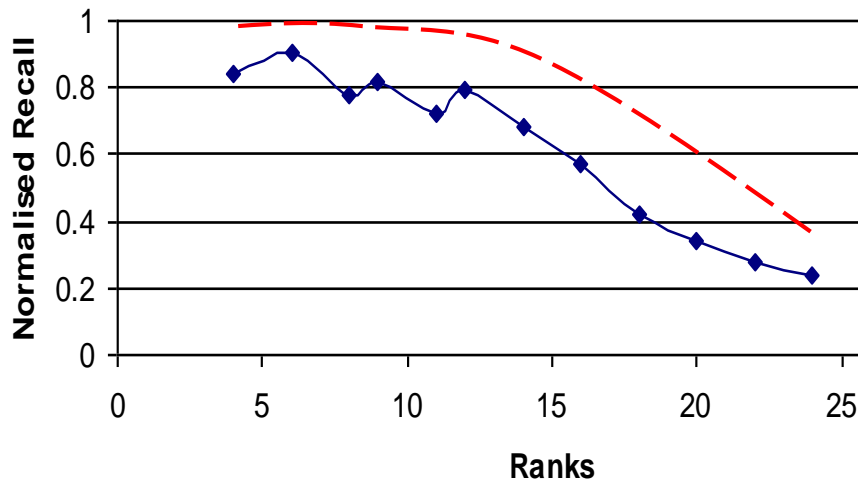
11. Mediators



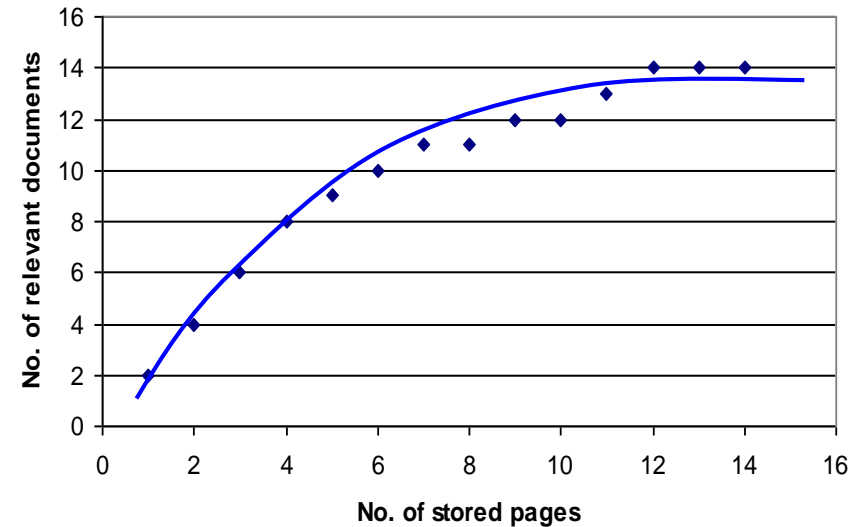
- ✓ Create/Clone/Manage the agents
- ✓ Translate the query
- ✓ Filter the relevant information
- ✓ Reformulate/Generate the query

12. Experimental results

Normalised precision curve



Feedback relevance



$$R_{norm} = 1 - \frac{\sum_{i=1}^n r_i - \sum_{i=1}^n i}{n(N - n)}$$

n - number of relevant documents
 r_i - rank of the i th document
 N - number of documents

13. Conclusions and future work

- ✓ SEA improve data access capabilities and communication ability
- ✓ The users are not aware by the context in which they search
- ✓ The necessity to deal with multiple profiles
- ✓ The need to express users' interest gradually



- ✓ To improve the query reformulation process
- ✓ Automate the profiling procedures
- ✓ Introduce some degrees of uncertainty in expressing the relevance.
- ✓ Extend our tool for collaborative searching



<http://airg.verena.ro/sea>