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

Cândea, 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

wittmann
& partner computer systems srl

AI research group

INES 2000, Portorož

Overview

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

wittmann
& partner computer systems srl

AI research group

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)

	Autonomy	Cooperation	Adaptation
Relevance Feedback		✧	✧
Information Extraction		✧	
Multimedia Retrieval		✧	
Effective Retrieval		✧	✧
Routing & Filtering	✧	✧	✧
Interfaces & Browsing		✧	✧
Term Expansion			
Efficiency & Flexibility	✧	✧	✧
Distributed IR	✧	✧	
Integrated solutions	✧	✧	✧

wittmann
& partner computer systems srl

AI research group

4. Current Trends
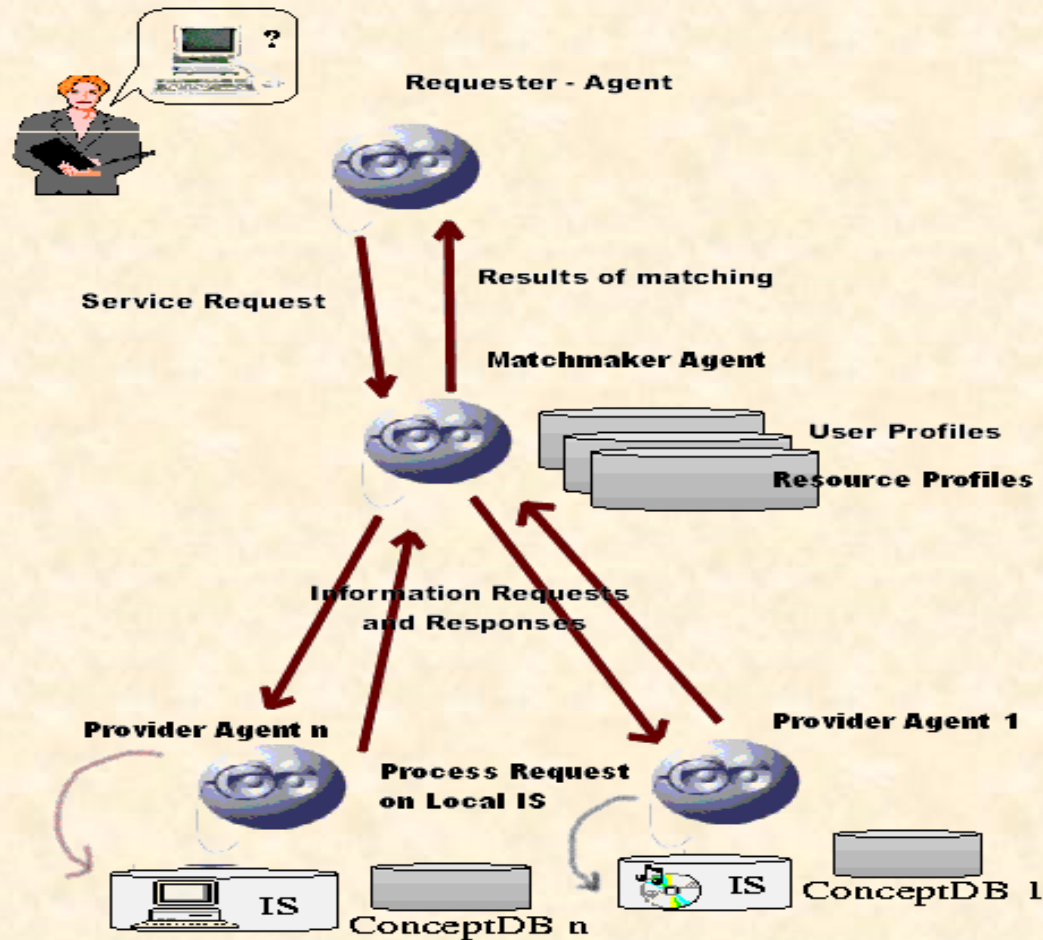
5. Design Issues

✖ 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*
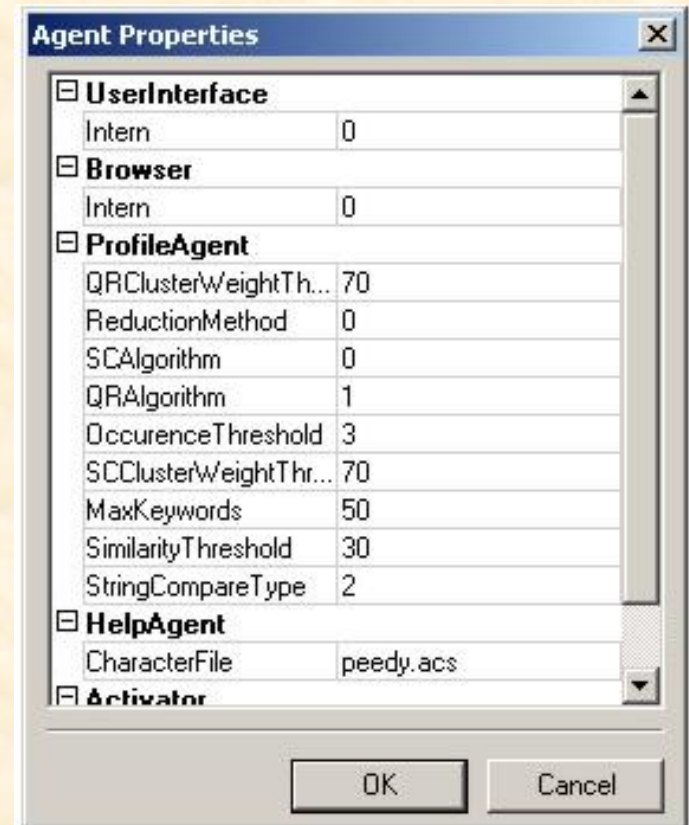 ✓ *content*

✖ To be easily embedded in any kind of application

6. SEA Architecture
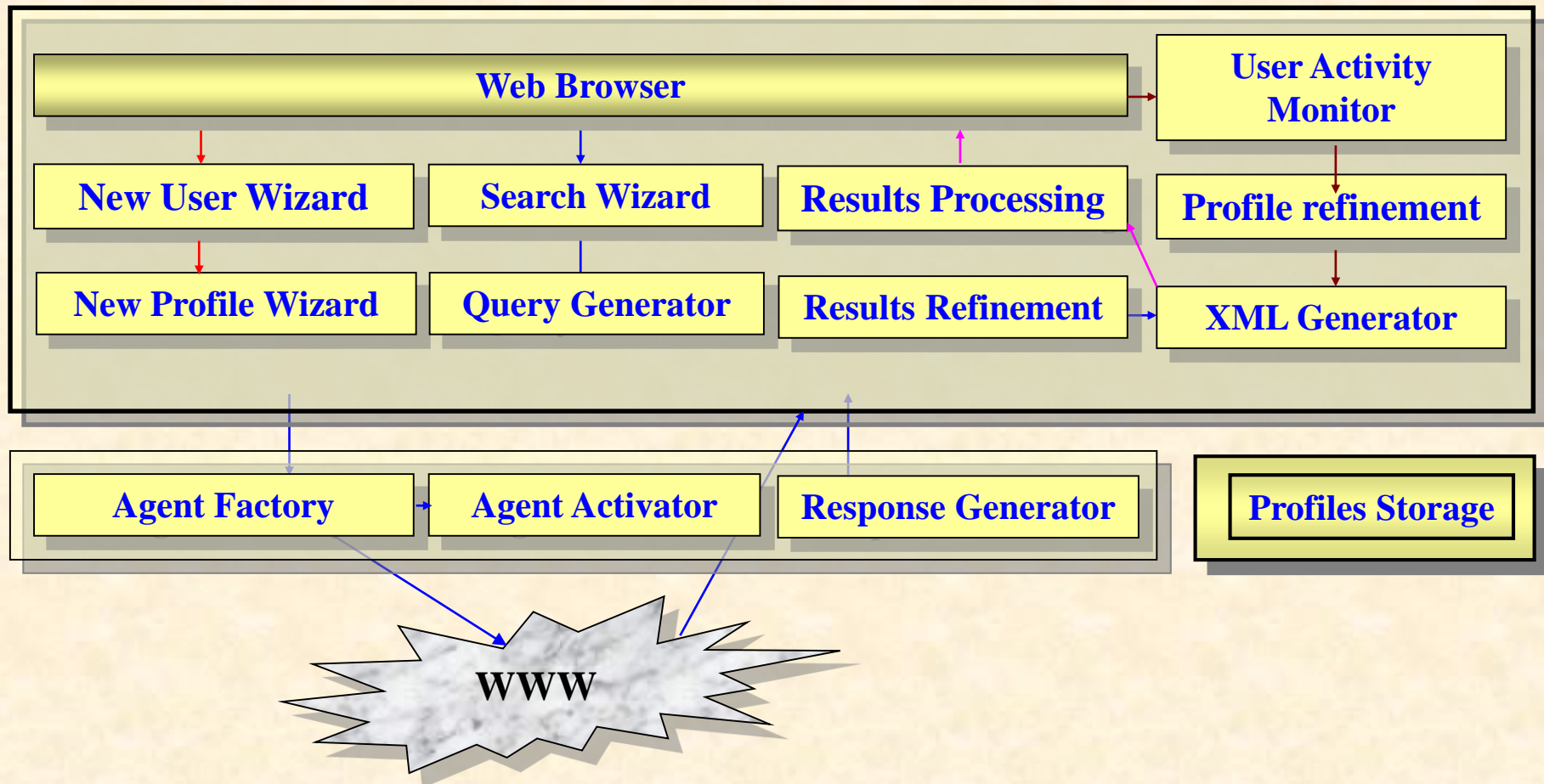
7. User Modeling

$$Sim\ (d_i,\ d_j) = 2\ X\ \big|\ d_i \cap d_j\ \big|\ /\ \big|\ d_i\ \big|\ +\ \big|\ d_j\ \big|$$

$$w_i = (0.5 + 0.5\ tf(i)\ /\ tf_{max})\ (log\ (n\ /\ df(i))\ )$$

$$p(w,\ m) = q(w)\ *\ m$$

$$u(w,\ m,\ s) = m + z(t)\ w$$

w	- a Web page
m	- user's interests
$tf(i)$	- the term frequency
$df(i)$	- 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

Agent Properties

⊟ **UserInterface**	
Intern	0
⊟ **Browser**	
Intern	0
⊟ **ProfileAgent**	
QRClusterWeightTh...	70
ReductionMethod	0
SCAlgorithm	0
QRAlgorithm	1
OccurenceThreshold	3
SCClusterWeightThr...	70
MaxKeywords	50
SimilarityThreshold	30
StringCompareType	2
⊟ **HelpAgent**	
CharacterFile	peedy.acs
⊟ **Activator**	

OK   Cancel

8. Interface Agent

Adaptability

Adaptivity

9. The Application

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>
```
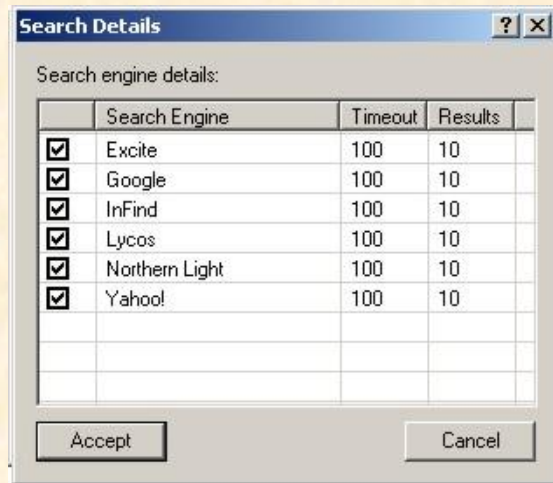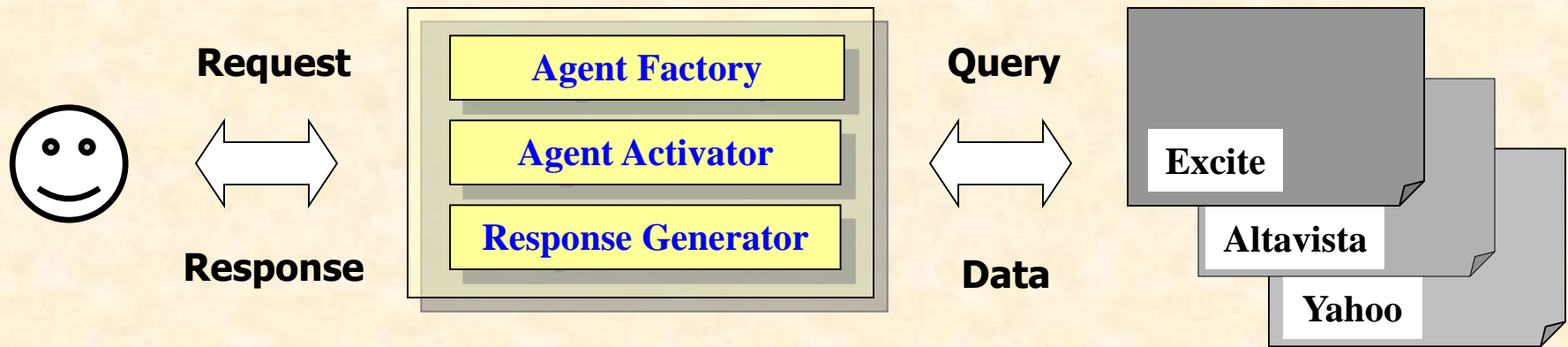
11. Mediators

Request

Agent Factory

Agent Activator

Response Generator

Response

Query

Data

Excite

Altavista

Yahoo

Search Details   ? X

Search engine details:

	Search Engine	Timeout	Results
☑	Excite	100	10
☑	Google	100	10
☑	InFind	100	10
☑	Lycos	100	10
☑	Northern Light	100	10
☑	Yahoo!	100	10

Accept   Cancel
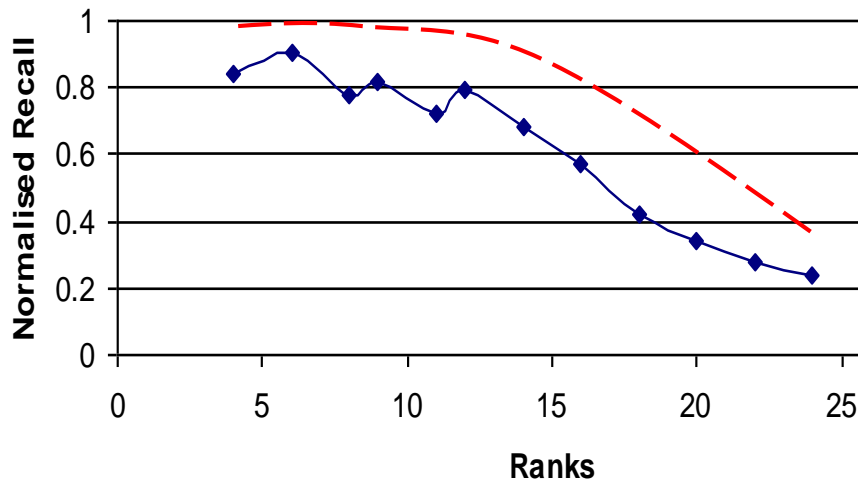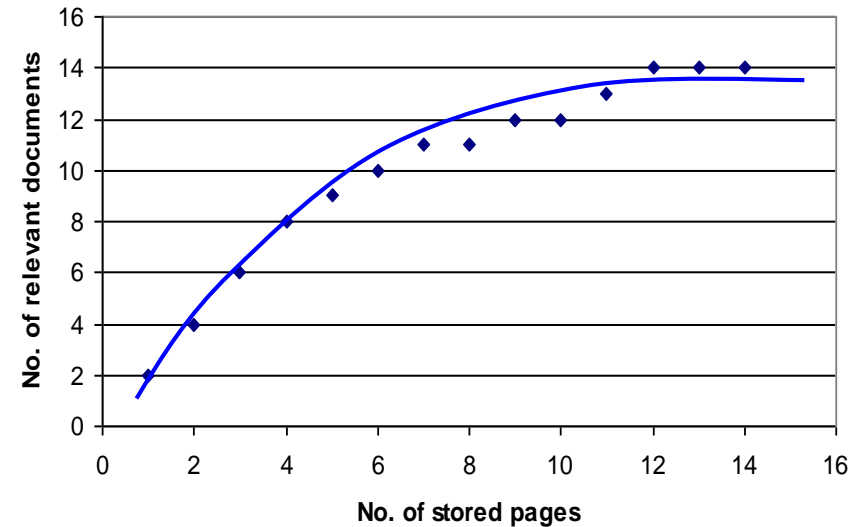
✓ Create/Clone/Manage the agents

✓ Translate the query

✓ Filter the relevant information

✓ Reformulate/Generate the query

wittmann
& partner computer systems srl

AJ research group

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

wittmann
& partner computer systems srl

AJ research group

http://airg.verena.ro/sea