

# Integrating information extraction agents into a tourism recommender system

HAIS'10



**Sergio Esparcia**  
Víctor Sánchez-Anguix  
Estefanía Argente  
Ana García-Fornes  
Vicente Julián

Grupo de Tecnología Informática  
**Inteligencia Artificial**

# OUTLINE

- Introduction
- Social-Net Tourism Recommender System
- Information agents add-on
  - Information extraction agents
  - Information classification agents
- Experiments
- Conclusions

# OUTLINE

- Introduction
- Social-Net Tourism Recommender System
- Information agents add-on
  - Information extraction agents
  - Information classification agents
- Experiments
- Conclusions

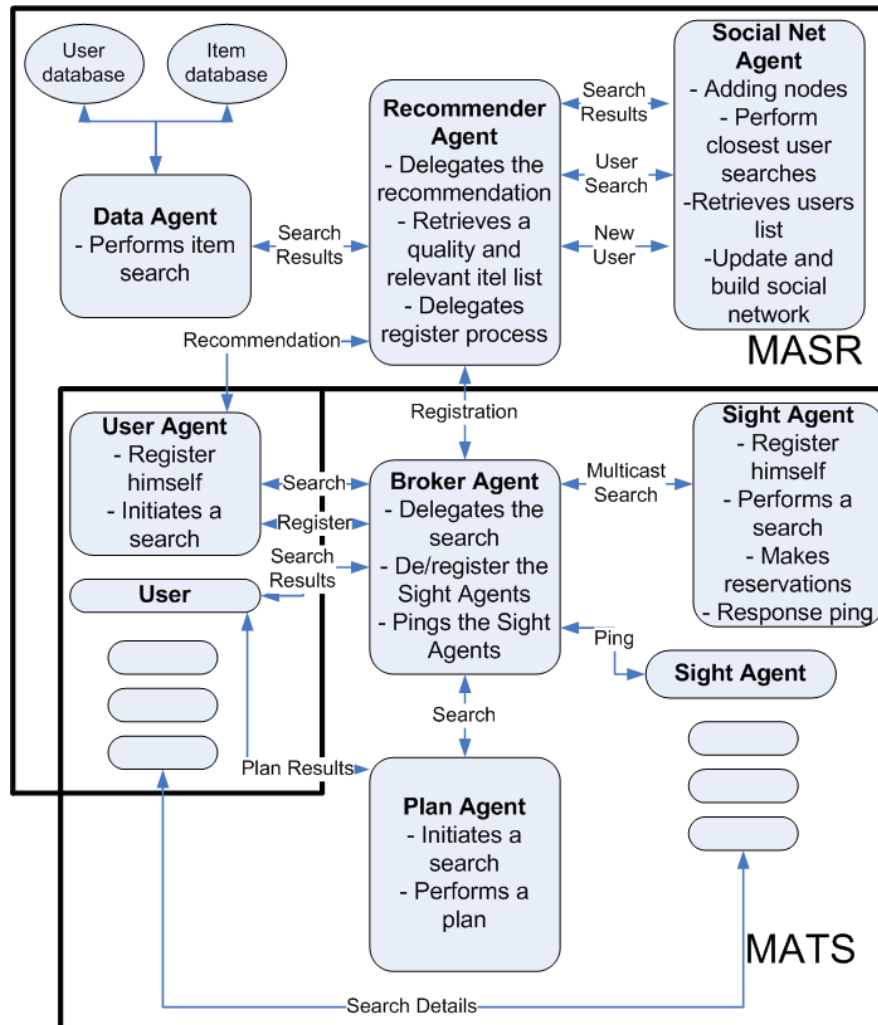
# INTRODUCTION

- Recommender systems: Tourism industry
  - Social-Net Recommender System (STRS)
- Problems
  - Keeping information up-to-date
  - Third parties are also offering services (enhance recommendations)
- Goals: Add-on for STRS
  - Keeping system user information up-to-date
  - Extract information from third parties to enhance recommendations

# OUTLINE

- Introduction
- Social-Net Tourism Recommender System
- Information agents add-on
  - Information extraction agents
  - Information classification agents
- Experiments
- Conclusions

# SOCIAL-NET RECOMMENDER SYSTEM



# OUTLINE

- Introduction
- Social-Net Tourism Recommender System
- Information agents add-on
  - Information extraction agents
  - Information classification agents
- Experiments
- Conclusions

# INFORMATION AGENTS ADD-ON

- Two different kinds of agents:
  - Information extraction agents
  - Information classification agents
- Information found on the Web
- Natural Language Processing (NLP) as classification mechanism



# INFORMATION EXTRACTION ADD-ON

- **Information extraction agents (IE):**
  - Wrapper architecture
    - HTML request
    - Analyze HTML structure: look for specific patterns
    - Extract relevant information of the event
  - Some information (category) needs to be inferred
    - Send events description to information classification (IC) agents
    - Wait for IC agents response

# INFORMATION EXTRACTION ADD-ON

- **Information classification agents (IC):**

- Each IC agent is specialized in classifying one event category
- Use NLP to categorize event descriptions
- Rule based system:
  - Term Strength rules

- Hyperonym rules  $\rightarrow SC_H(w_i) = \frac{|S(w_i)| - (i-1)}{\sum_{k=1}^i |S(w_i)|_k}$

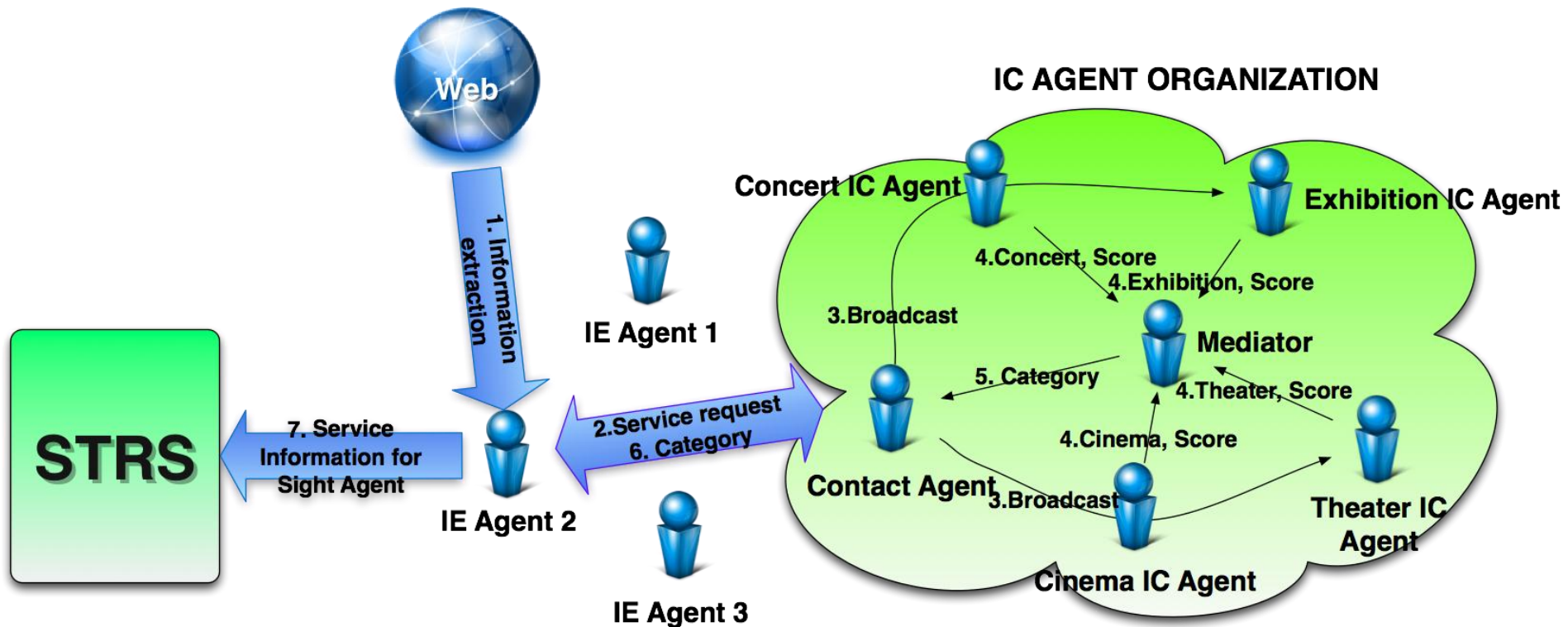
# INFORMATION EXTRACTION ADD-ON

- **Information classification agents (IC):**
  - Score for a word  $w_i$  = score of the matching rule that maximizes the score vote for  $w_i$
  - Final score vote for a service description = sum of the final scores produced by all the words

# INFORMATION EXTRACTION ADD-ON

- **Information classification agents (IC):**
  - All IC agents form a mediated agent organization
  - Contact agents: Classification service
  - Mediator agent:
    - Decides final category
    - Adjusts voting power ( $vp$ ) (e.g. past experiences)
    - $Category(W) = Expertise(\operatorname{argmax}_{a_i \in ICS} vp_{a_i} * SC_{a_i}(W))$

# INFORMATION EXTRACTION ADD-ON



# OUTLINE

- Introduction
- Social-Net Tourism Recommender System
- Information agents add-on
  - Information extraction agents
  - Information classification agents
- Experiments
- Conclusions

# EXPERIMENTS

- First experiment:
  - Test classification accuracy
    - Three categories: Music, theater, exhibition
  - 600 balanced service description
    - 30% test, 70% training
    - $v p_{a_i} = 1$
    - Baseline: Term Strength

Method	Classification error (test)	Classification error (training)
Proposed method	11,11%	11,79%
Term Strength	16,67%	17,65%

# EXPERIMENTS

- Second experiment:
  - Test voting power:
    - 3 bad designed agents and the 3 previous agents
    - Voting power adjustment based on past classifications (updated each 10 service calls)
    - Adjustment rule:

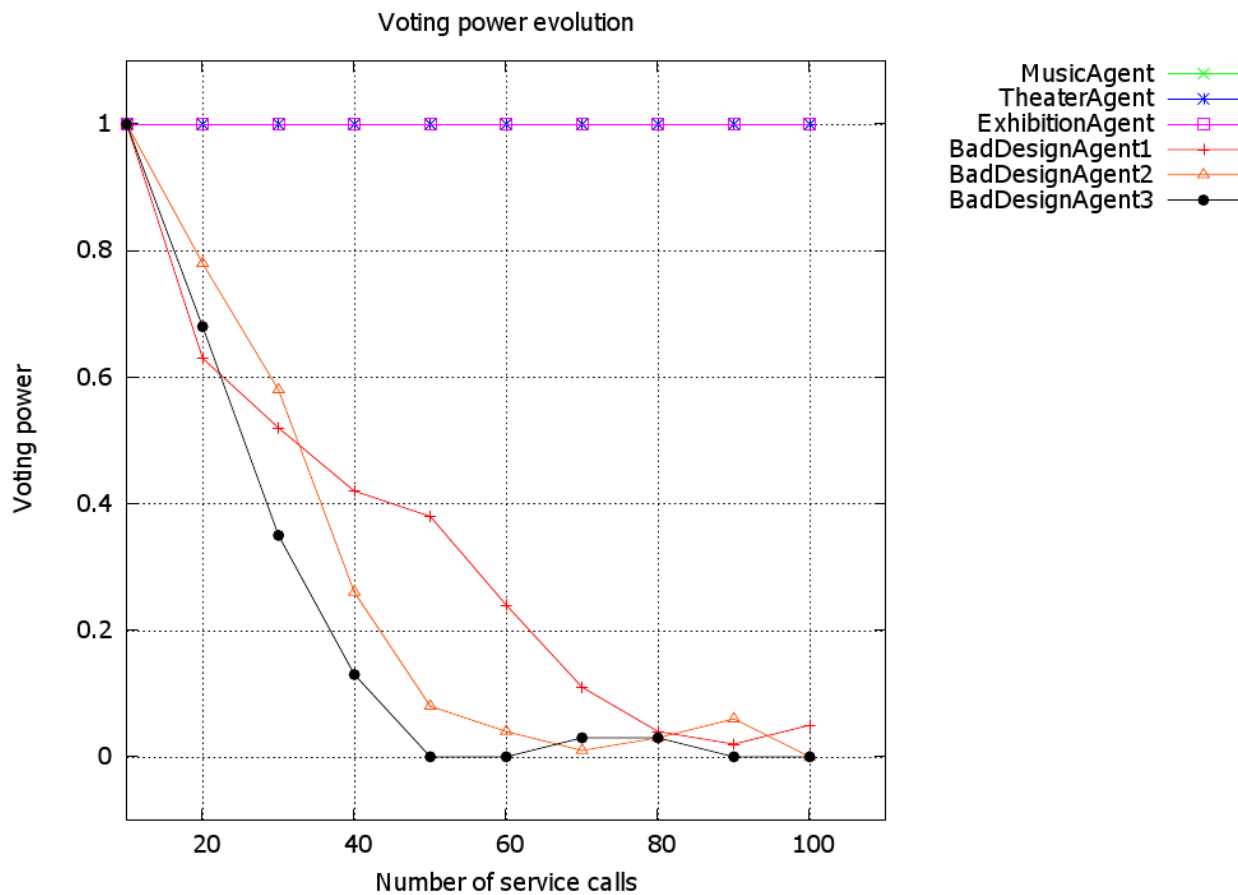
$$vp_{a_i}^{t+1} = vp_{a_i}^t - \frac{FP_{ai}}{|N_{other}|} + \frac{TP_{ai}}{|N|}$$

where:

- $vp_{a_i}^{t+1}$  is the new voting power
- $vp_{a_i}^t$  is the voting power of agent  $a_i$  in the last check
- $FP_{ai}$  is the number of false positives given by  $a_i$
- $TP_{ai}$  is the number of true positives given by  $a_i$
- $|N|$  is the total number (10) of service calls
- $|N_{other}|$  is the total number of service calls whose associated service category is not the one that  $a_i$  represents



# EXPERIMENTS



# OUTLINE

- Introduction
- Social-Net Tourism Recommender System
- Information agents add-on
  - Information extraction agents
  - Information classification agents
- Experiments
- Conclusions

# CONCLUSIONS

- An add-on for STRS was presented
  - Keeps information up-to-date
  - Retrieves information from third-parties
- Information extraction agents employ NLP to extract and classify information
- Bad designed agents can be neutralized by means of mediated voting processes

# THANKS FOR YOUR ATTENTION



## Any questions?

