Semantic Reasoning: A Path to New Possibilities of Personalization

Yolanda Blanco Fernández

yolanda@det.uvigo.es
University of Vigo (Spain)

5th European Semantic Web Conference
Tenerife, June 2008

Motivation

- Overload of information → Digital Revolution
- Recommender systems
  - Database
  - Users' profiles → preferences or needs
  - Recommendation strategies
    - Content-based filtering
    - Collaborative filtering

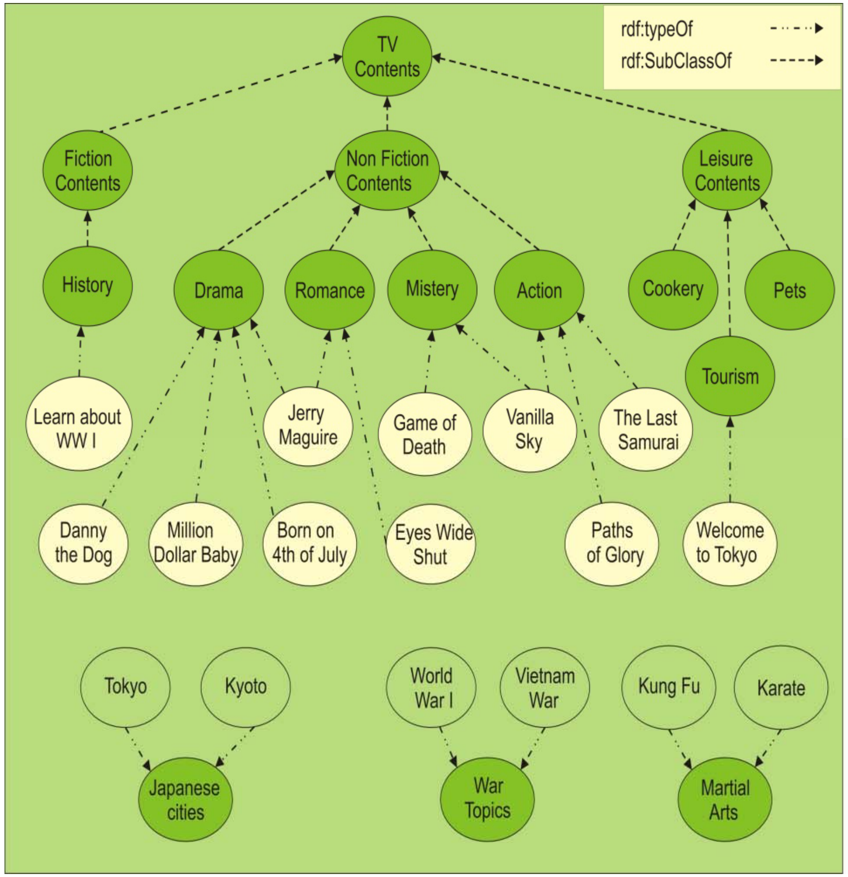
Recommendation Strategies

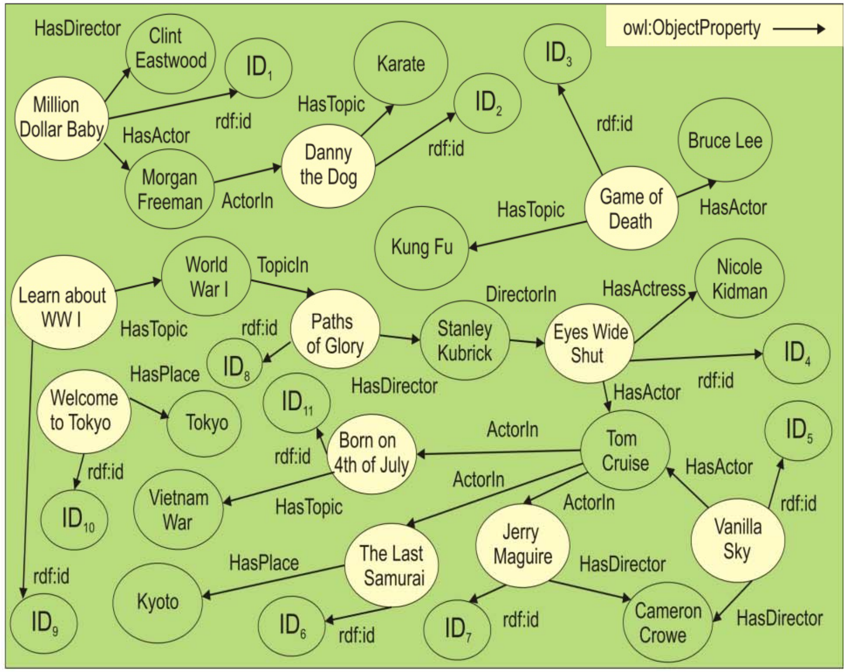
- Content-based filtering:
  - To suggest items similar to those defined in the user's profile → content-descriptions (attributes)
  - Syntactic matching techniques
  - Overspecialized recommendations
- Collaborative filtering:
  - To suggest items interesting for other users with similar preferences
  - Diverse recommendations, but other limitations:
    - Sparsity problem, privacy concerns…

Our Content-based Strategy

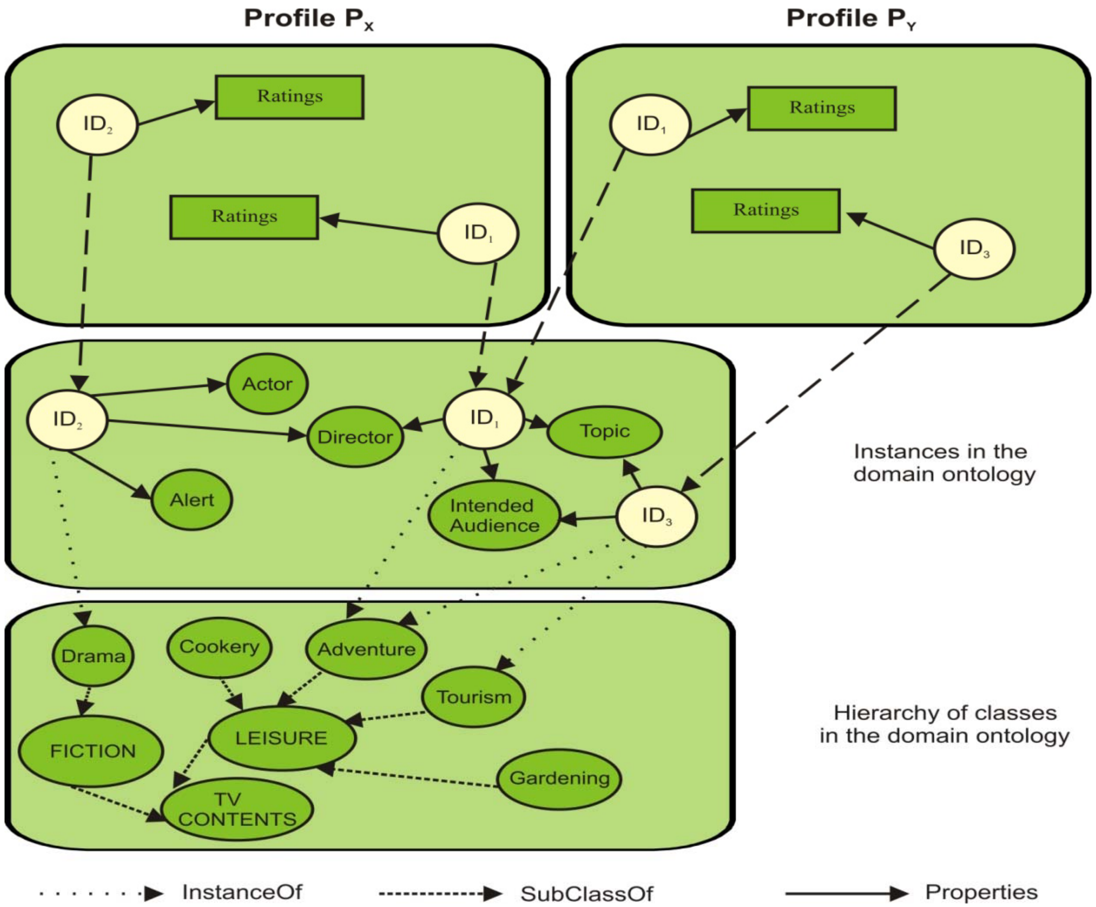
- To harness advantages and mitigate weaknesses of traditional content-based filtering:
  - Other users' preferences not necessary → privacy
  - Reasoning techniques → diversify recommendations
    - **Semantic Associations**
    - Spreading Activation techniques (**SA techniques**)
- Adapt reasoning techniques to meet personalization requirements of recommender systems.
- Reasoning framework must include: domain **ontology** and **user modeling technique**.

An Example of TV Ontology





User Modeling Technique



Our Reasoning-based Strategy

- Content-based filtering → To suggest items semantically related to the user's positive preferences.
- Two-phase strategy:
  - Filtering phase: Selects excerpts from ontology containing instances relevant for user, and infers semantic associations between specific items and user's preferences.
  - Recommendation phase: Processes inferred knowledge by SA techniques → detect concepts strongly related to user's preferences → **enhanced content-based recommendations**.

Filtering Phase: How do we find instances relevant for the user?

- First, the items defined in the user's profile are located in the ontology.

- Properties from these items are successively traversed, reaching new nodes:
  - If node is relevant → continue traversing its properties.
  - Otherwise → disregard the properties linking the reached node to others in the ontology.

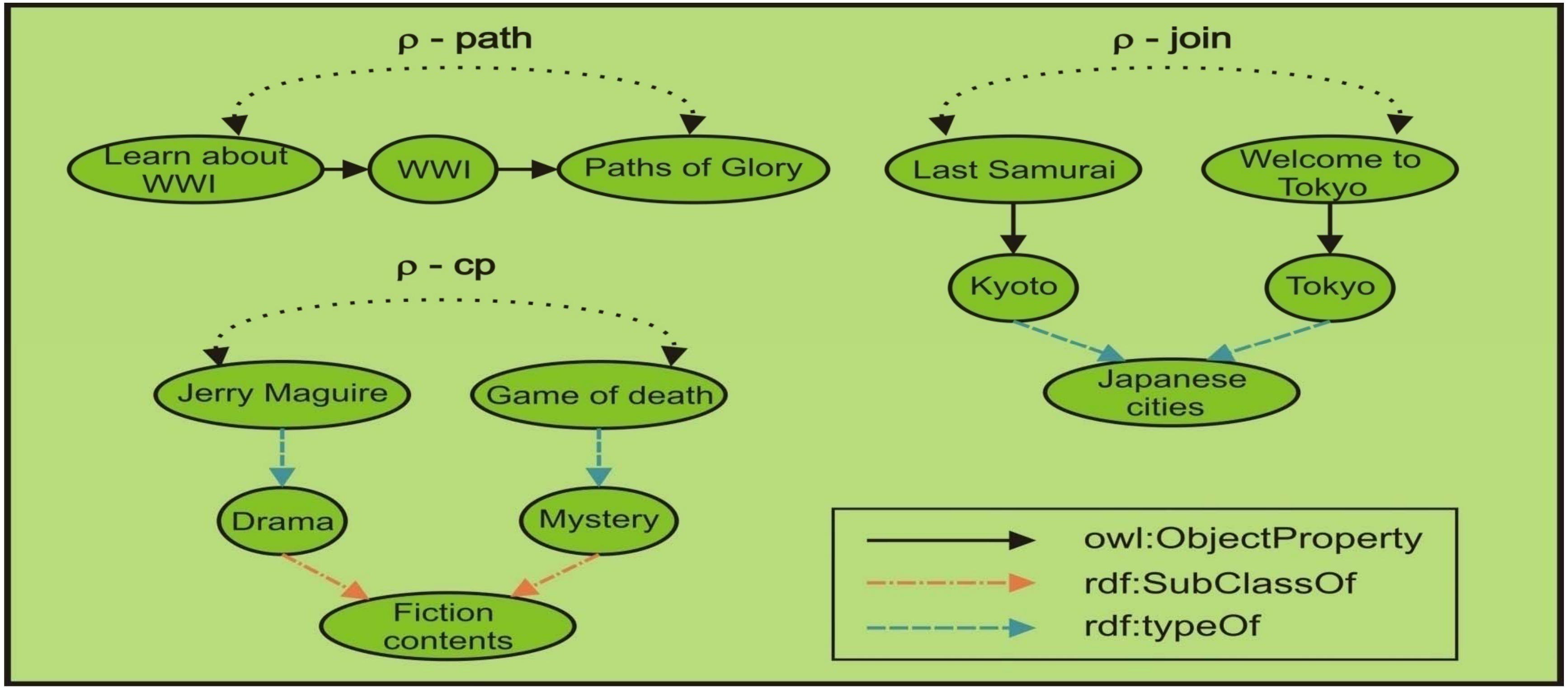
- Only instances of interest for the user are explored!

Filtering Phase: How do we compute the relevance of a node?

- The stronger the relationship between a node **N** and the user's preferences, the higher the relevance of **N**.
- Relevance value is measured by ontology-dependent filtering criteria:
  1. Length of chain of properties established between **N** and class instances in the user's profile:
     - The lower number of intermediate items, the more relevant **N**
  2. Hierarchical relationships between N and user's preferences.
  3. Implicit relationships detected by graph theory concepts:
     - High ***betweenness*** among **N** and class instances defined in the user's profile → **N** is strongly related to his preferences.

Filtering Phase: How do we infer Semantic Associations between items?

- Research project SemDis (Anyanwu and Sheth)



Recommendation Phase

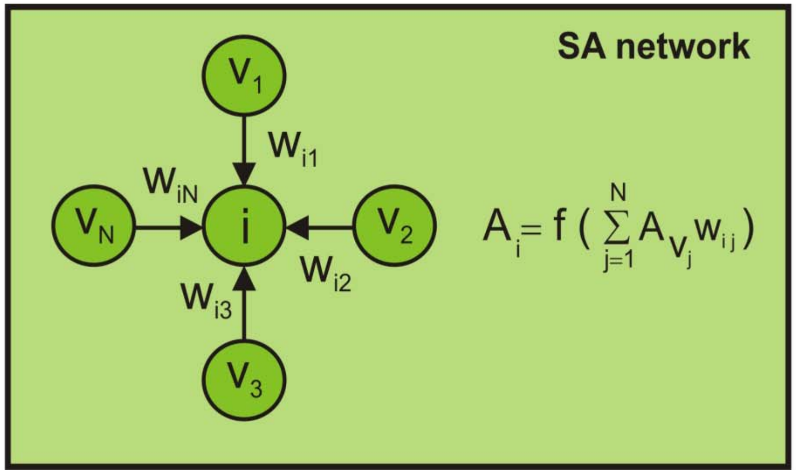
- Knowledge available after filtering phase:
  - Class and properties instances.
  - Semantic Associations between specific items.

- This network is processed by SA techniques → SA network:
  - Explore efficiently relationships among nodes interconnected in SA network.
  - Detect items strongly related to user's positive preferences → content-based recommendations

How do traditional SA techniques work?

- Exploration of huge knowledge networks:
  - Nodes → activation level (relevance of the node in the network)
  - Links → static weights (strength of relationships between linked nodes)

Selection of initially activated nodes → **Spreading of activation levels** → **Nodes strongly related to those initially activated**



Recommendation Phase: How do we create the user's SA network?

- Nodes → Class instances selected by filtering phase.
- Links → Property instances and semantic associations.

- How do we weight the links of the user's SA network?:
  - Traditional static weights are not valid for recommender systems due to personalization requirements.
  - The links are weighted according to the user's preferences:
    - The stronger the relationship between the two linked nodes and the user's preferences, the higher the weight of the link.
    - Weights of links are **updated as the user's preferences change over time.**

How do we select our content-based recommendations?

- Nodes initially activated → items in the user's profile.
- Initial activation levels → ratings

- After spreading process…
  - Items with highest activation levels are suggested to the user.
  - Strongly related to his preferences → High quality content-based recommendations.
  - Items are ranked acccording to their activation levels.

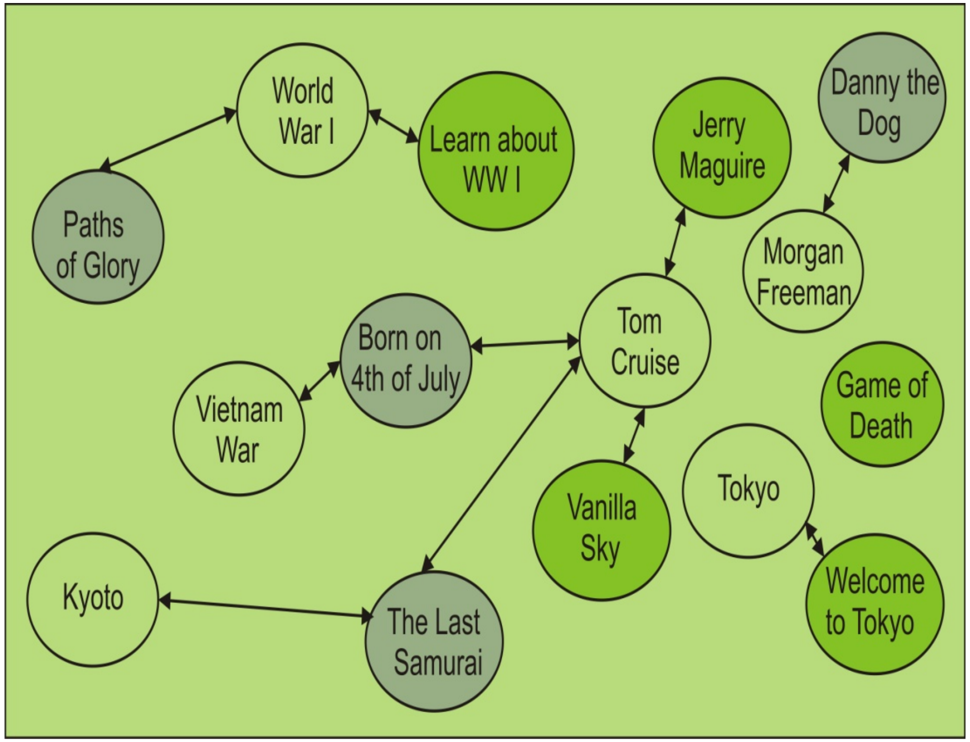
A Sample Scenario

- Digital TV domain → overload of audiovisual contents and interactive applications.
- Select content-based recommendations for Mary → TV ontology

Mary's positive preferences	**Mary's negative preferences**
- Wellcome to Tokyo - Learn about World War I - Vanilla Sky - Jerry Maguire	- Million Dollar Baby (*Morgan Freeman*) - Game of death (*martial arts*)

Filtering Phase: Selecting instances relevant for Mary

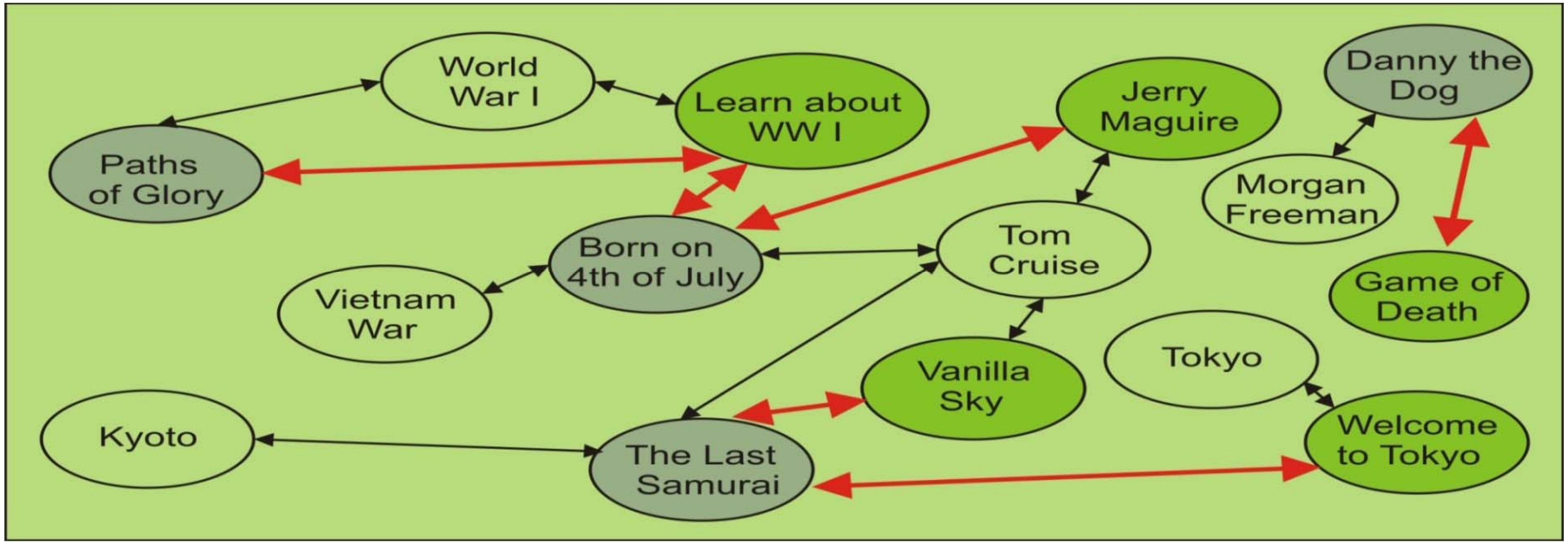
- Born on 4th July – Jerry Maguire: Drama movies
- The Last Samurai – Vanilla Sky: Action movies
- Vietnam War – World War I: War topic
- Tokyo – Kyoto: Japanese cities
- Danny the Dog – Million dollar baby: Morgan Freeman
- Danny the Dog – Game of death: Martial arts



Filtering Phase: Inferring Semantic Associations between TV programs

Semantic Associations	Why are they inferred?
➢ ρ-path (Jerry Maguire, Born on 4th of July)	Tom Cruise
➢ ρ-join (Welcome to Tokyo, The last samurai)	Japanese cities
➢ ρ-join (Learn about WWI, Born on 4th of July)	War topic
➢ ρ-cp (Vanilla Sky, The last samurai)	Action contents
➢ ρ-join (Danny the Dog, Game of death)	Martial arts
➢ ρ-path (Danny the Dog, Million dollar baby)	Morgan Freeman

Recommendations Phase: Suggesting TV programs to Mary



- Our strategy suggests…
  - Paths of glory
  - Born on the 4th of July
  - The last samurai
- Our strategy does not suggest…
  - Danny the Dog

Experimental Evaluation: Setting

- 400 undergraduate students from University of Vigo
- TV ontology with programs extracted from BBC web site and Internet Movie DataBase
- Users rated 400 programs in the range [-1,1]

- We evaluated our reasoning-based strategy against:
  - O'Sullivan *et al.* → content-based filtering and association rules to measure similarity between programs.
  - Mobasher *et al.* → semantics-enhanced collaborative filtering

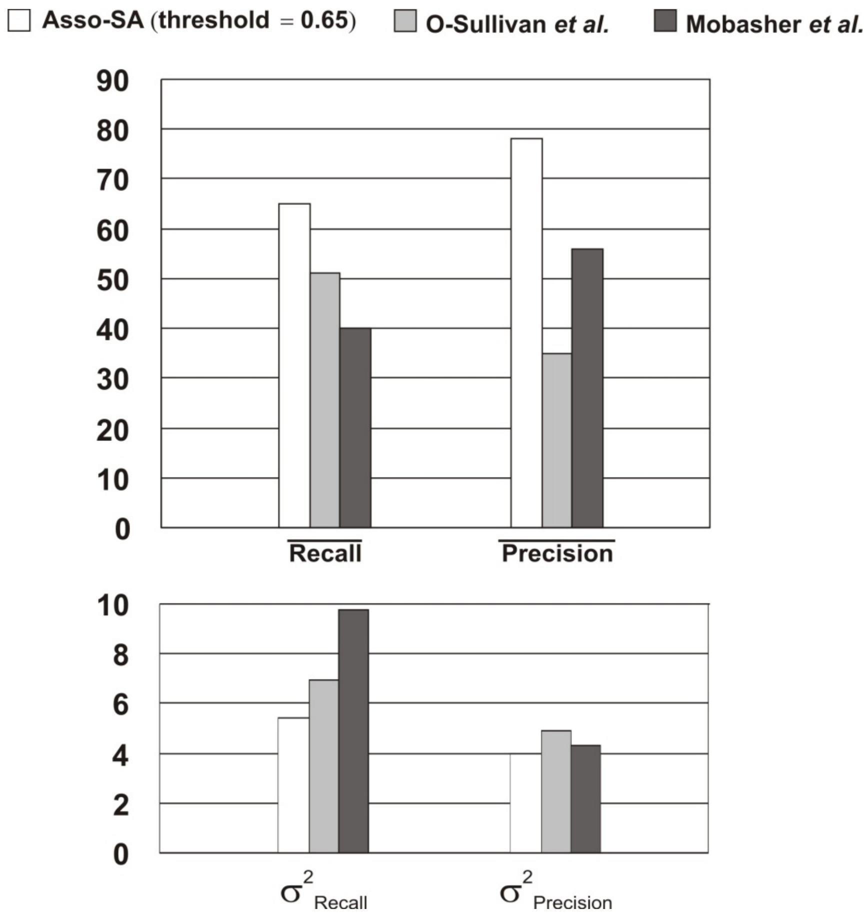
Experimental Evaluation: Setting

- **Training profiles** (160 users) → compute values needed in the strategies devoid of our reasoning capabilities.
- **Test profiles** (240 users) → execute 3 evaluated strategies:
  - 20 programs to initialize the test users' profiles → great sparsity level
  - 380 programs and ratings to measure recommendation accuracy → evaluation data

- Recall: percentage of interesting programs that were suggested.
- Precision: percentage of programs suggested that are appealing to the user.
- Average and variance of recall and precision over 240 tests users.

Experimental Evaluation: Results

- Semantic reasoning leads to highest recall and precision values.
- Low overlap between programs defined in test users:
  - O'Sullivan *et al.*→ difficult to detect association rules between programs, and measure similarity between programs.
  - Mobasher *et al.* → difficult to detect neighbors and offer collaborative recommendations.



Conclusions

- Content-based strategy enhanced by reasoning:
  - Semantic associations
  - SA techniques
- Diverse recommendations → items semantically related to the user’s preferences → beyond syntactic matching
- Positive and negative preferences are considered.
- Recommendations adapted as user’s preferences evolve.
- Flexible enough to be used in multiple domains.
- Significant increases in recall and precision w.r.t. reasoning-devoid strategies.

Further Work

- Automatic adjustment of thresholds:
  - Filtering phase
  - Recommendation phase
  - Dependent on domain ontology and user feedback.

- New experiments with subscribers of the cable network of Spanish operator R (*http://www.mundo-r.com*).

Thank you for your attention!