Identifying Users Stereotypes with Semantic Web Mining

José Palazzo Moreira de Oliveira

Sandro José Rigo

Instituto de Informática – Universidade Federal do Rio Grande do Sul (UFRGS)
{palazzo,sjrigo}@inf.ufrgs.br

Contents

- **Introduction and motivation**
- **Related works**
- **Semantic Web Mining**
- **Developed architecture**
- **Experiments**
- **Conclusions**

Introduction and motivation

- **Web cognitive overload**
    - ◦ Huge documents number
    - ◦ Different Web systems
    - ◦ User characteristics diversity

- **Improve user experience**
    - ◦ Personalized results
    - ◦ General steps:
        - • User profile acquisition
        - • Adaptation rules
        - • Interface modifications

Introduction and motivation

- **Web Usage Mining**
  - User profile acquisition
  - Frequent access patterns mining

- **Problems:**
  - Patterns generated without semantic information
  - User identification needed

User profile acquisition

- **Diversity of related works**
  - Observation of query terms and navigation sequence (Dummais, 2006)
  - LIP based ontology (Razmerita, 2004)
  - Cognitive aspects and usage information (Souto, 2005)
  - Indication of interests and preferences (Sah e Hall, 2007)

Main Objectives

○ Users classes identification

○ Short term tasks

○ Integration:

- Web usage information
- Semantic Information (application domain)

Related works

- **Some related works:**
  - Web Usage Mining, clustering (Baraglia e Silvestri, 2007)
  - Web Usage Mining, semantic annotation, association rules (Eirinaki, 2006)
  - Domain Ontologies, clusters (Mobasher, 2005)
  - Frequent access patterns mining, taxonomies (Esposito, 2004)
  - Clustering, similarity relations (Vieira, 2005)
  - Ontology acquisition (Staab, 2000; Stumme, 2002)
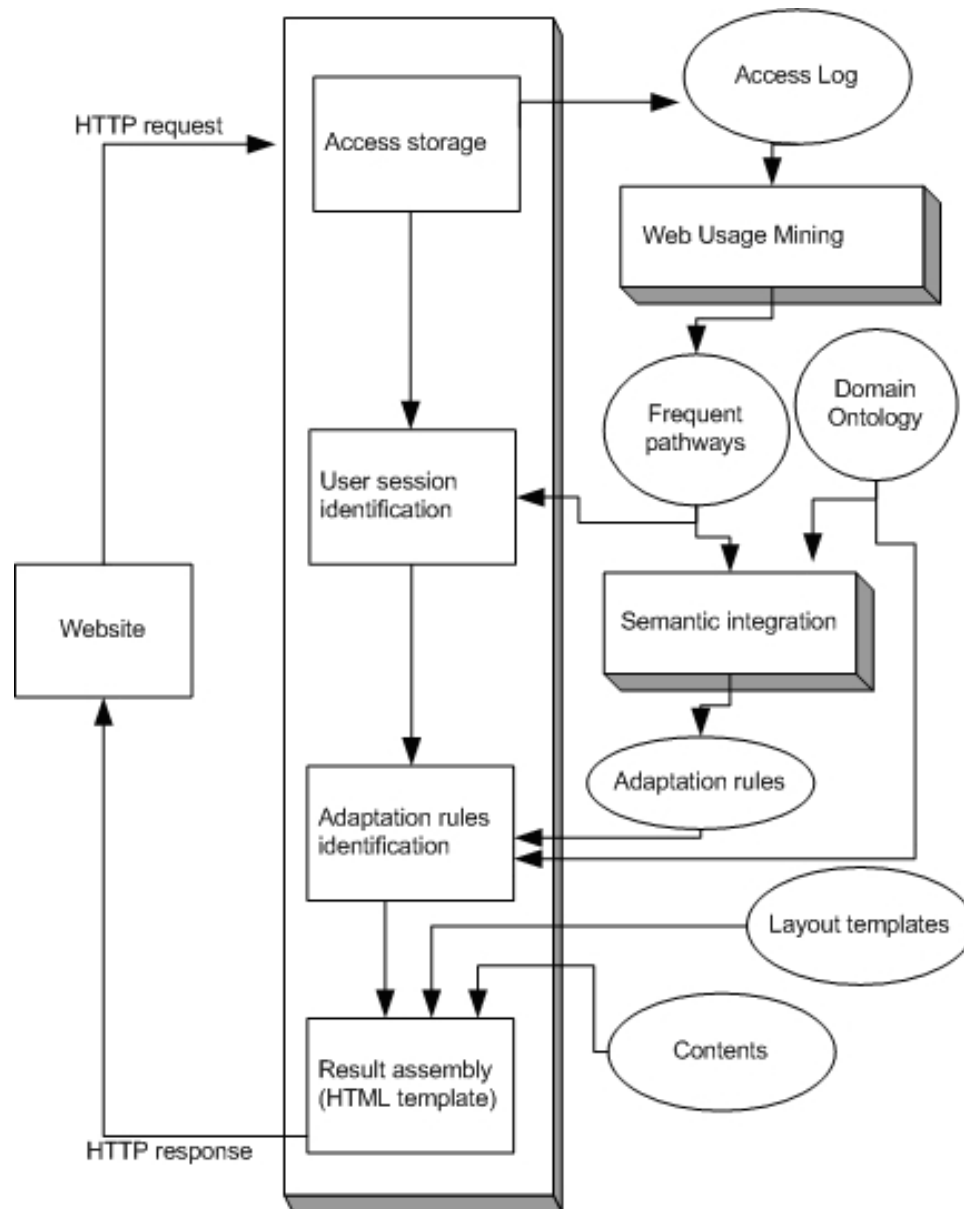- **General approaches :**
  - Ontology/taxonomy acquisition
  - Integration of semantic information
    - patterns generation
    - adaptation

Semantic web mining approach

- Frequent access patterns generation
- Ontology with web site structure and other relations
  - Content type
  - Elements precedence
- Semantic integration with patterns
  - Ontology queries
  - Relations between frequent access patterns
- Patterns differentiation and similarities identification
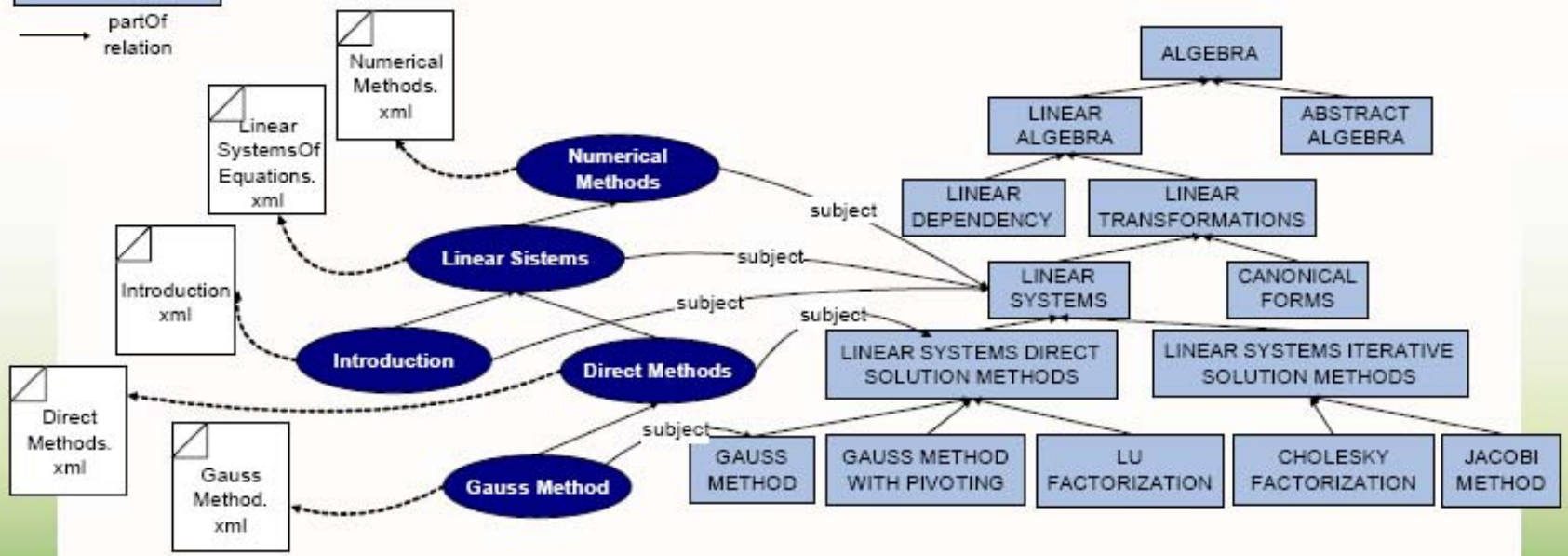- Adaptation resources

HTTP request

Access storage

Access Log

Web Usage Mining

User session identification

Frequent pathways

Domain Ontology

Website

Semantic integration

Adaptation rules

Adaptation rules identification

Layout templates

Result assembly (HTML template)
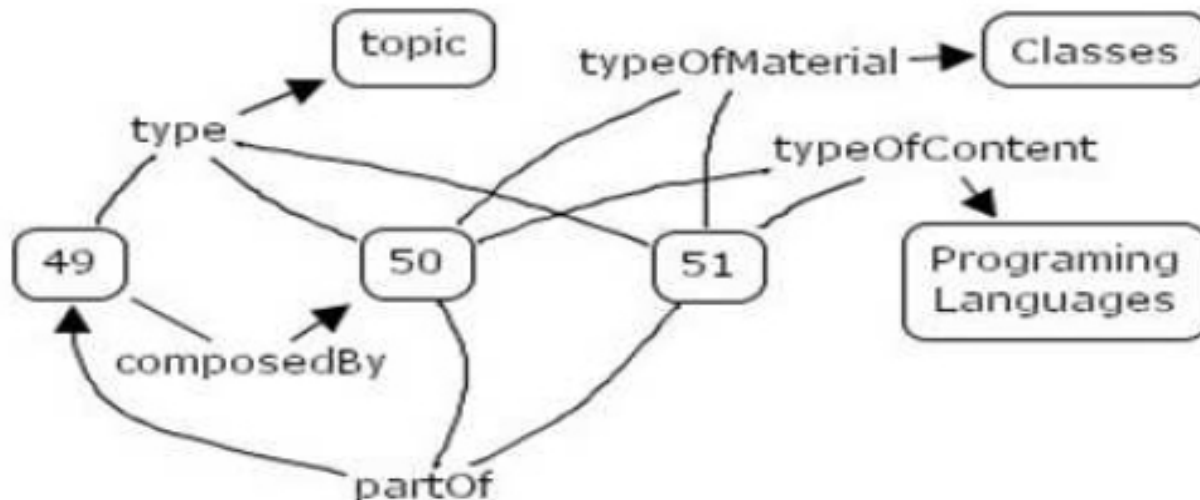
Contents

HTTP response

Hyperspace

Knowledge Space

Semantic contexts

- **General view**

a) Pageviews,   Pages IDs
   3,   49, 50, 51

b) Sequence: 2 IDs: 49, 50 (composed_by_topic) [type - topic]
     Sequence: 2 IDs: 50, 51 [type - topic] [partOf - ID_49]
              [type_of_material - TM_Classes]
   [type_of_content - Programming_language]

N	Patterns	Semantic contexts
1	13, 12, 33	(composedBy)(requisiteOf)[type-topic][typeOfMaterial-TM_ORGANIZATION] (composedBy) [type-topic] [typeOfMaterial-TM_ORGANIZATION]
2	12, 33, 44	(composedBy) [type - topic] [typeOfMaterial - TM_ORGANIZATION] (composedBy) [type - topic] [typeOfMaterial - TM_ORGANIZATION]
3	49, 50, 51	(composedBy) [type - topic] [type - topic] [partOf - ID_49] [typeOfMaterial - TM_CLASS] [typeOfContent- linguagem_de_programacao]
4	20, 46, 47, 48	(composedBy) [type - topic] [typeOfMaterial - TM_CLASS] [type - topic] [partOf - ID_20] [typeOfMaterial - TM_CLASS] [typeOfContent- algoritmo] [type - topic] [partOf - ID_20] [typeOfMaterial - TM_CLASS] [typeOfContent- linguagem_de_programacao]
5	49, 50, 51, 52	(composedBy) [type - topic] [type - topic] [partOf - ID_49] [typeOfMaterial - TM_CLASS] [typeOfContent- linguagem_de_programacao] [type - topic] [partOf - ID_49] [typeOfMaterial - TM_CLASS] [typeOfContent- linguagem_de_programacao]
6	62, 64, 67, 72	(requisiteOf) [partOf - ID_13] [typeOfMaterial - TM_ORGANIZATION] (requisiteOf) [partOf - ID_13] [typeOfMaterial - TM_ORGANIZATION] (requisiteOf) [partOf - ID_13] [typeOfMaterial - TM_ORGANIZATION] [requisiteOf - ID_78]

Semantic contexts
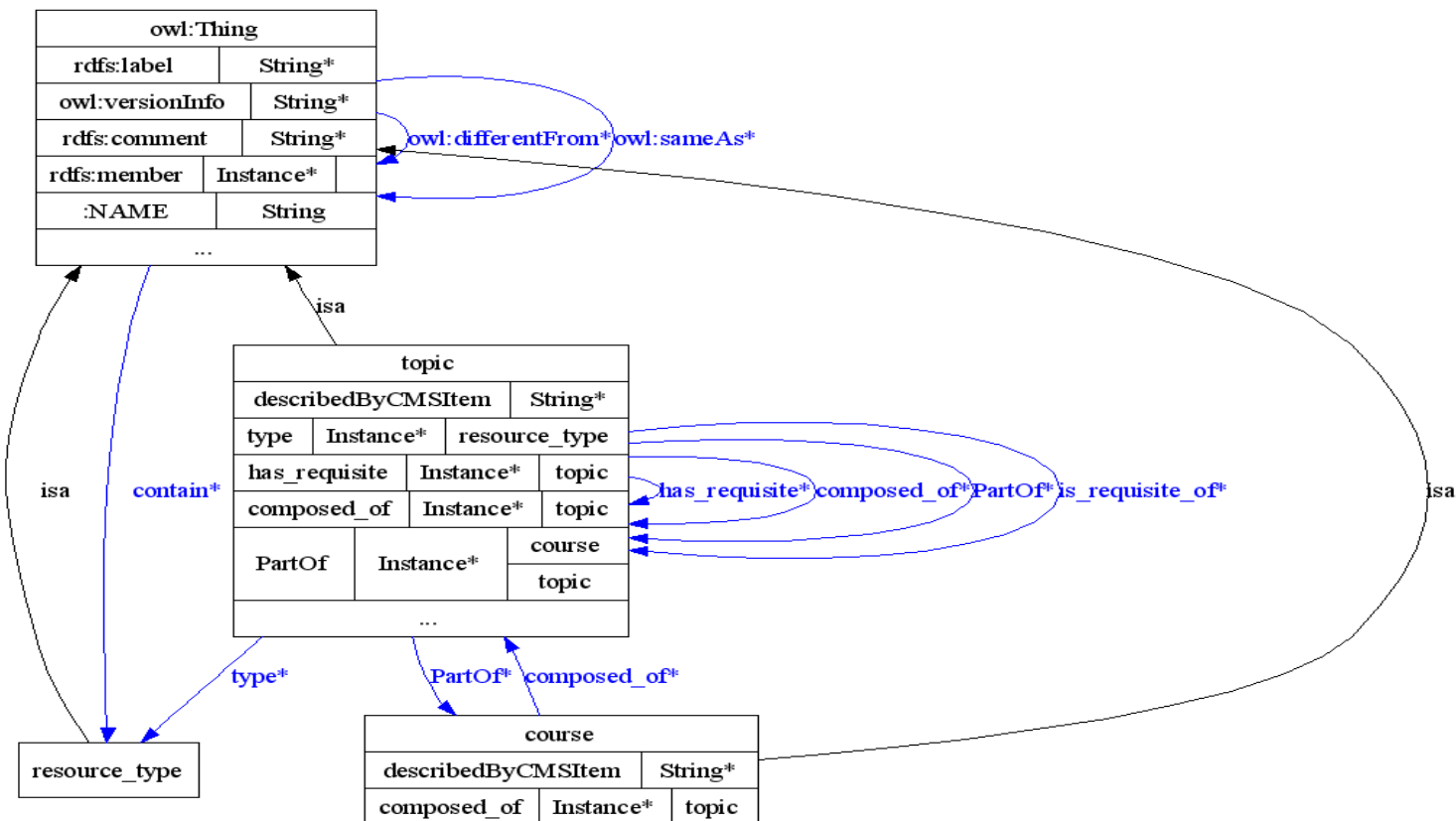
- **Advantages and possibilities**
  - ◦ Navigation differentiated contexts
  - ◦ Frequent access patterns similarities
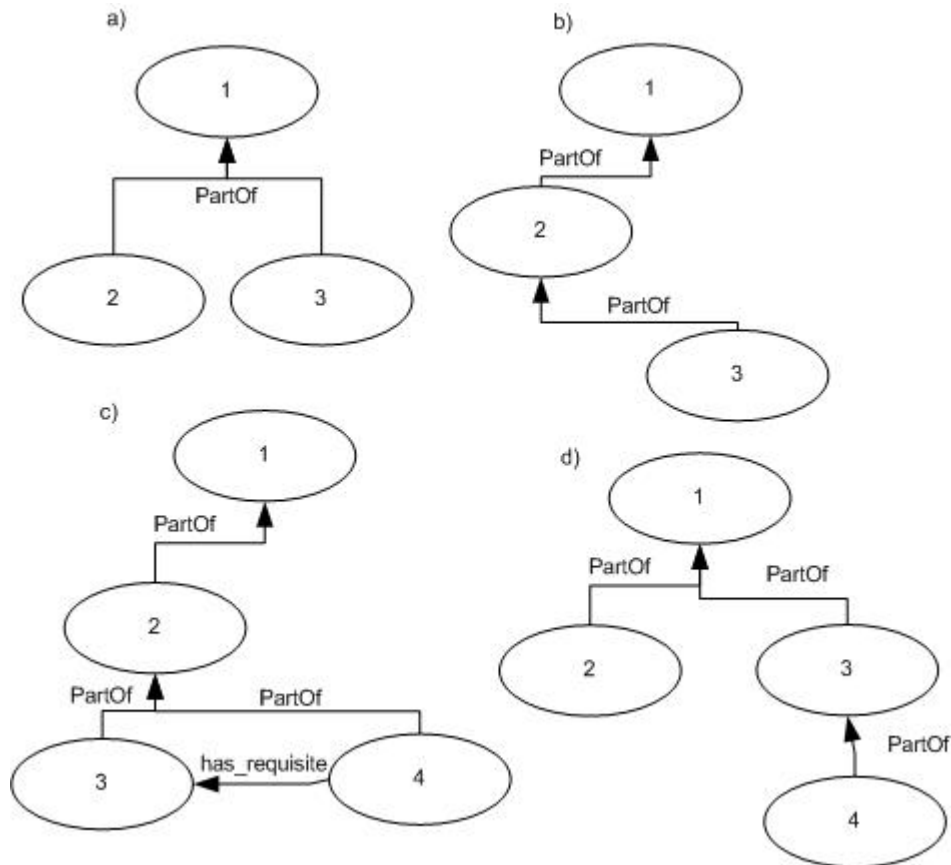
Semantic annotation and ontology

- **Domain Ontology**

  - Identification of relevant relations

  - Domain concepts

  - Web site pages annotation

Adaptation

- **Adaptation rules possibilities:**

  - User session partial coincidence

  - User session and semantic context similarity

  - Ontology descriptions

    - Complementariety

    - Composition

Adaptation

- **Adaptation examples:**

Developed architecture

- **Steps:**

  - Usage data acquisition
  - Frequent access patterns
  - Domain ontology description
  - Web site semantic annotation
  - Semantic integration
  - Adaptation rules
  - Adaptation generation
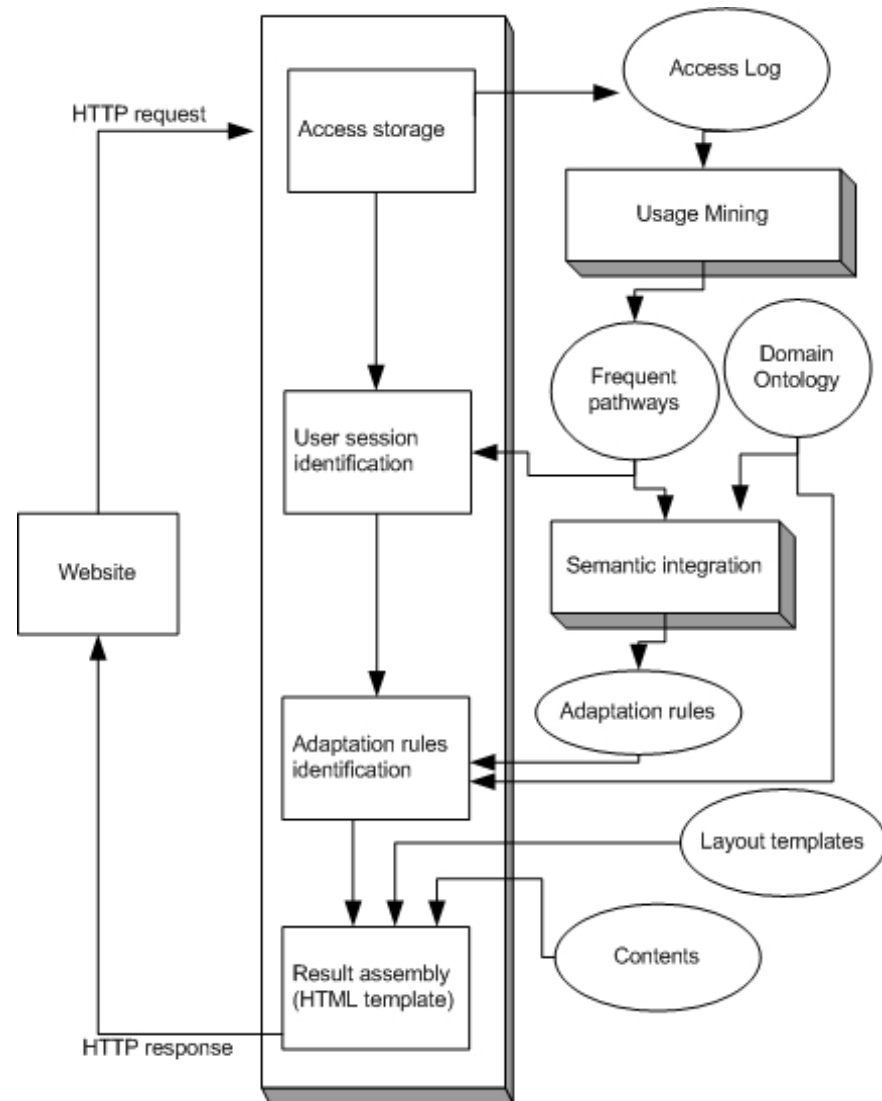
Experiments

- Experiment was carried out over a period of ten months

- Adaptive hypermedia application target to an programming language discipline

- Material available to students, with adaptations being generated

- The suggested adaptations are monitored

- Access to adaptations is compared to the normal Website items

Conclusions

- Approach for the acquisition of user stereotypes based on Web Usage Mining and domain ontology

- Adaptive hypermedia application

- Semantic information and usage information integration

- Discovering navigational behavior related to tasks executed in a Website

- The proposed method can generate valuable information by relating the Web usage data and the semantic information.

Future work

- Applying web page internal elements in the generation of patterns

- Study of integration possibilities with content information

- Experiments in applications other than educational

- Insertion of new relations in the domain ontology

Thank you.

Identifying Users Stereotypes with Semantic Web Mining

José Palazzo Moreira de Oliveira

Sandro José Rigo

Instituto de Informática – Universidade Federal do Rio Grande do Sul (UFRGS)
{palazzo,sjrigo}@inf.ufrgs.br