Service Retrieval Based on Behavioral Specification and Quality Requirements

Daniela Grigori, Veronika Peralta, Mokrane Bouzeghoub

PRISM, University of Versailles, France

Motivation

- Service retrieval:
  - Software components, scientific applications, composite web services, process models
  - A retrieval process step: matchmaking based on behavioral specification (process model) and quality requirements

Motivating scenarios

- Web services integration
  - Finding services that have compatible business protocols (response time, availability, security)
- Scientific applications
  - Finding similar scientific workflows (data accuracy, completeness)
- Composite web services discovery
  - The discovery process selects the most similar models form the library and propose them for reuse by extension or modification (response time)

Goal

- Developing matching techniques allowing:
  - delivery of partial matches
  - evaluation of semantic distance between these matches and user requirements
  - compose partial matches to satisfy user requirements

Approach

- Reduce the problem of service behavioral matching to an adorned graph matching problem
  - Graph represents the internal process of the service
  - Adornment represents quality constraints
- User requirements and process models are transformed to graphs
  - Apply the error correcting graph matching algorithm
  - Extended by a quality evaluation procedure

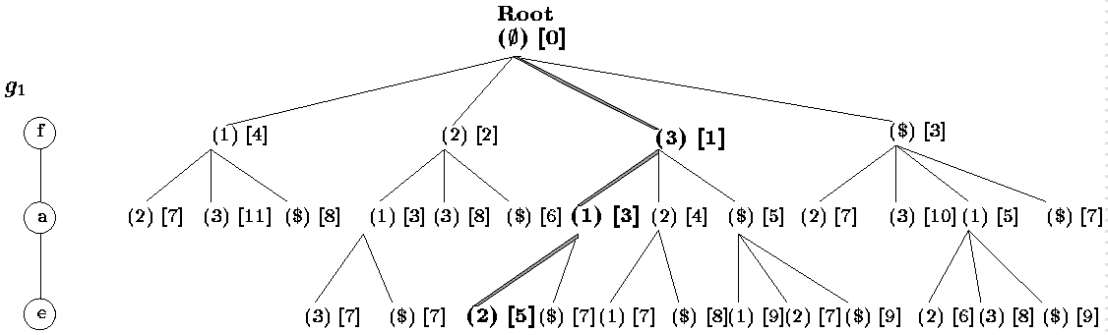
Error correcting graph matching

[Messmer, Bunke]

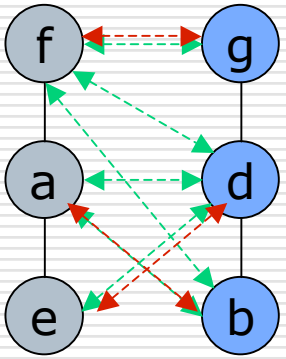
- Definition : Ec subgraph isomorphism
  - Given two graphs G and G', an error correcting (ec) subgraph isomorphism f from G to G' is a 2-tuple $f=(\Delta, f_\Delta)$ where
    - Δ is a sequence of edit operations such that there exists a subgraph isomorphism from $\Delta(G)$ to G'.
    - f is a subgraph isomorphism from $\Delta(G)$ to G'.
- Graph edit operations:
  - Substituting labels
  - Deleting vertices
  - Inserting and suppressing edges

Ecgm Algorithm - Example

- similar to A* [Shapiro, Haralick]



Mapping



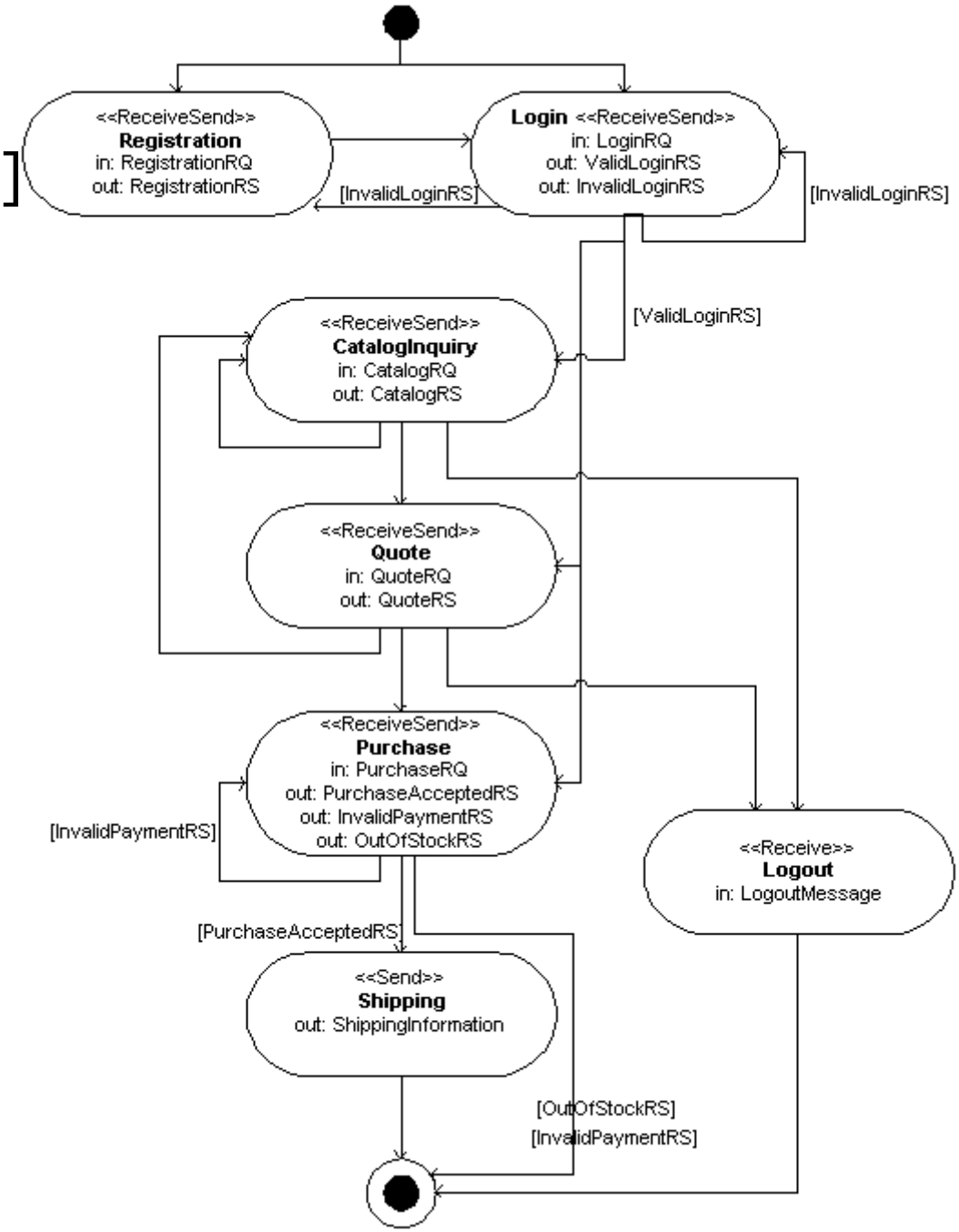
g1 g2

Behavioral matchmaking

- Conversation protocol (Ex: WSCL) -> graph

WSCL	**Graph**
Interactions (model the actions of the conversation as document exchanges between two participants).	Vertices
Transitions (specify the ordering relationships between interactions)	Edges

WSCL example [WSCL]



Cost function for graph edit operations

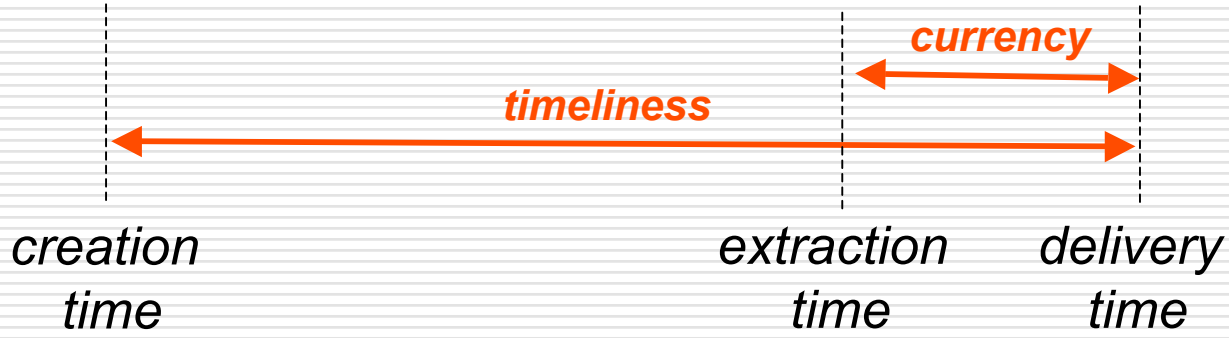
- Cost for deletion/insertion of an edge and vertex can be set to constants
- Cost for substituting a label :
  -  labels have meaningful names
    - Cost is based on the measure of the linguistic similarity between two concepts based on their names [METEOR] (uses various names and string matching algorithms like Ngram, synonym matching, abbreviation expansion, tokenization)
  -  labels are annotated with ontological concepts
    - The cost function is the distance between the concepts in the ontology (ex : Wu and Palmer similarity measure)

Non-functional properties

- Information quality is characterized by attributes which help to rank:
  - Data delivered to users (e.g., freshness, accuracy, completeness)
  - Processes that manipulate this data (response time, reliability, availability, security)
- The relevance of the service retrieval can be enhanced by taking into account quality factors
  - Adorn process graphs with quality features
  - Estimate the quality feature for the service and compare with user quality constraints

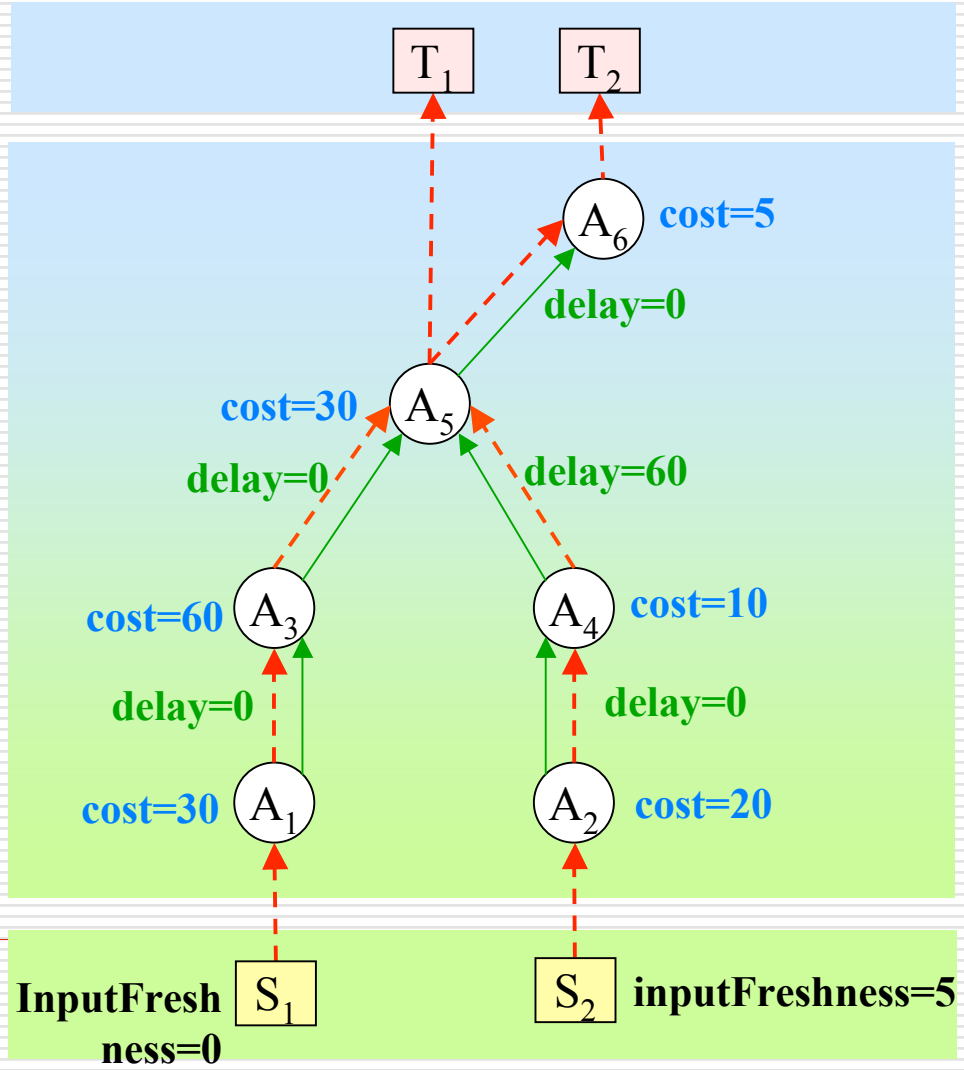
Data Freshness

- Data freshness quality factors:
  - Currency: Gap extraction – delivery
    - How stale is data with respect to sources?
  - Timeliness: Gap creation/update – delivery
    - How old is data? Is its age appropriate?



Quality Evaluation Approach

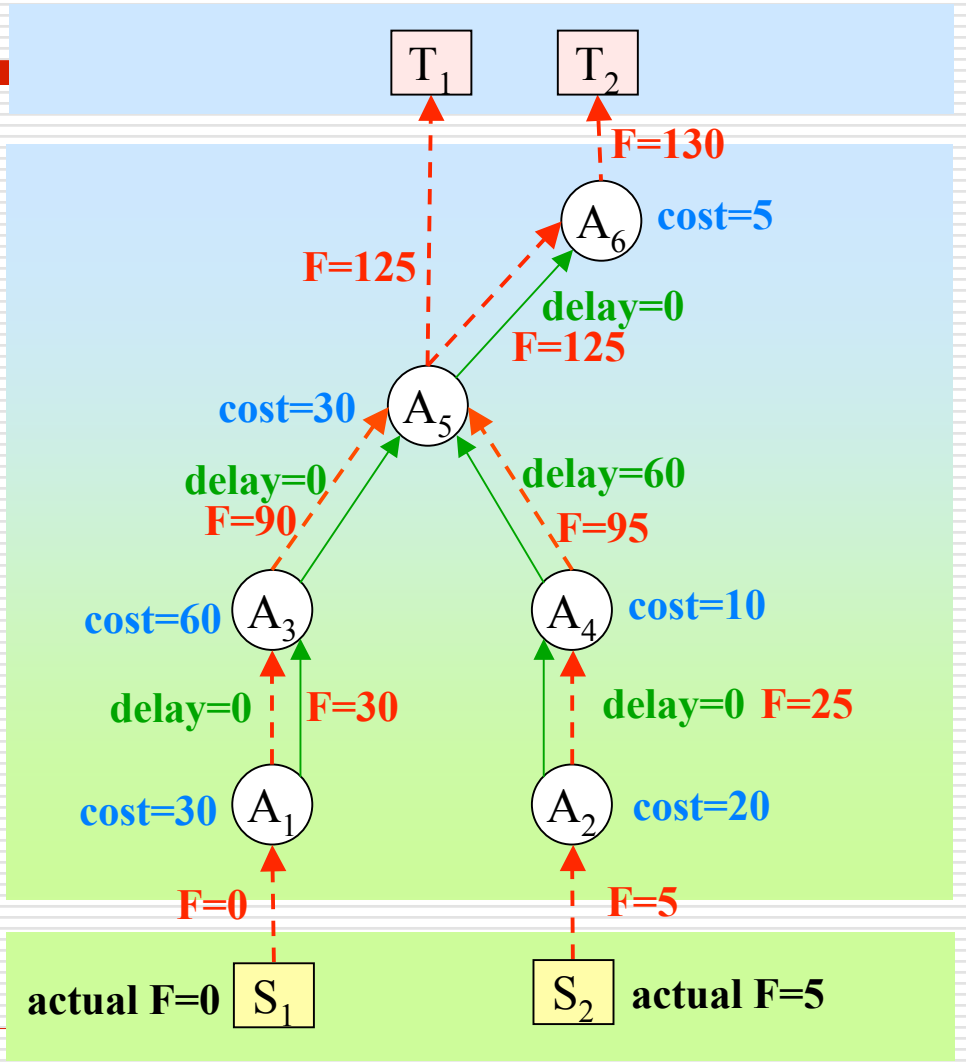
- Adornments of graphs
  - Activity processing cost
  - Synchronization delay
  - Input data actual freshness
- Quality evaluation as graph traversal methods



Freshness Evaluation Algorithm

Algorithm principle:

- **Source nodes S:**
  - **Freshness (S)** = ActualFreshness (S)
- **Other nodes N:**
  - **Freshness (N)** = Freshness (P) + Delay (P,N) + Cost (N)
- **Freshness of several input nodes are combined**



Using quality requirements in Behavioral Service retrieval

- Matching under Quality Requirements
  - Quality requirements = quality thresholds that catalog graphs must verify
- Matching with Quality Expectations
  - Quality requirements express expectations

Matching under Quality Requirements

- Retrieval steps:
  - Match the library graphs
  - Evaluate data freshness and eliminate the candidates that do not achieve freshness requirements;
  - Rank the candidate graphs according to their structural similarity and retrieve the best or the k best.

Matching with Quality Expectations

- Retrieval steps :
  - Match the library graphs in order to obtain the isomorphic subgraphs and calculate the structural similarity measure;
  - Evaluate data freshness and calculate the qualitative similarity measure;
    - SQ = (ExpectedFreshness – ActualFreshness) / ExpectedFreshness
  - Rank the candidate graphs according to their global similarity and retrieve the best or the k best.
    - $S = \alpha \times SS + \beta \times SQ$

Conclusion and ongoing work

- A solution for service retrieval based on behavioral specification and quality requirements
- Prototypes for behavioral matchmaking and freshness evaluation
- Ongoing work :
  - Integration of the two functionalities
  - Extending the quality evaluation procedure to other quality factors (response time, data accuracy)
  - Applying constraints and heuristics to cut down the computational effort of the matchmaking algorithm to a manageable size