

Semantic trace comparison at multiple levels of abstraction

Stefania Montani^a, Manuel Striani^b, Silvana Quaglini^c, Anna Cavallini^d, Giorgio Leonardi^a

^a DISIT, Computer Science Institute, University of Piemonte Orientale, Alessandria, Italy; ^b Department of Computer Science, University of Torino, Italy; ^c Department of Electrical, Computer and Biomedical Engineering, University of Pavia, Italy; ^d Istituto di Ricovero e Cura a Carattere Scientifico Fondazione C. Mondino, Pavia, Italy - on behalf of the Stroke Unit Network (SUN) collaborating centers

1. Process trace comparison

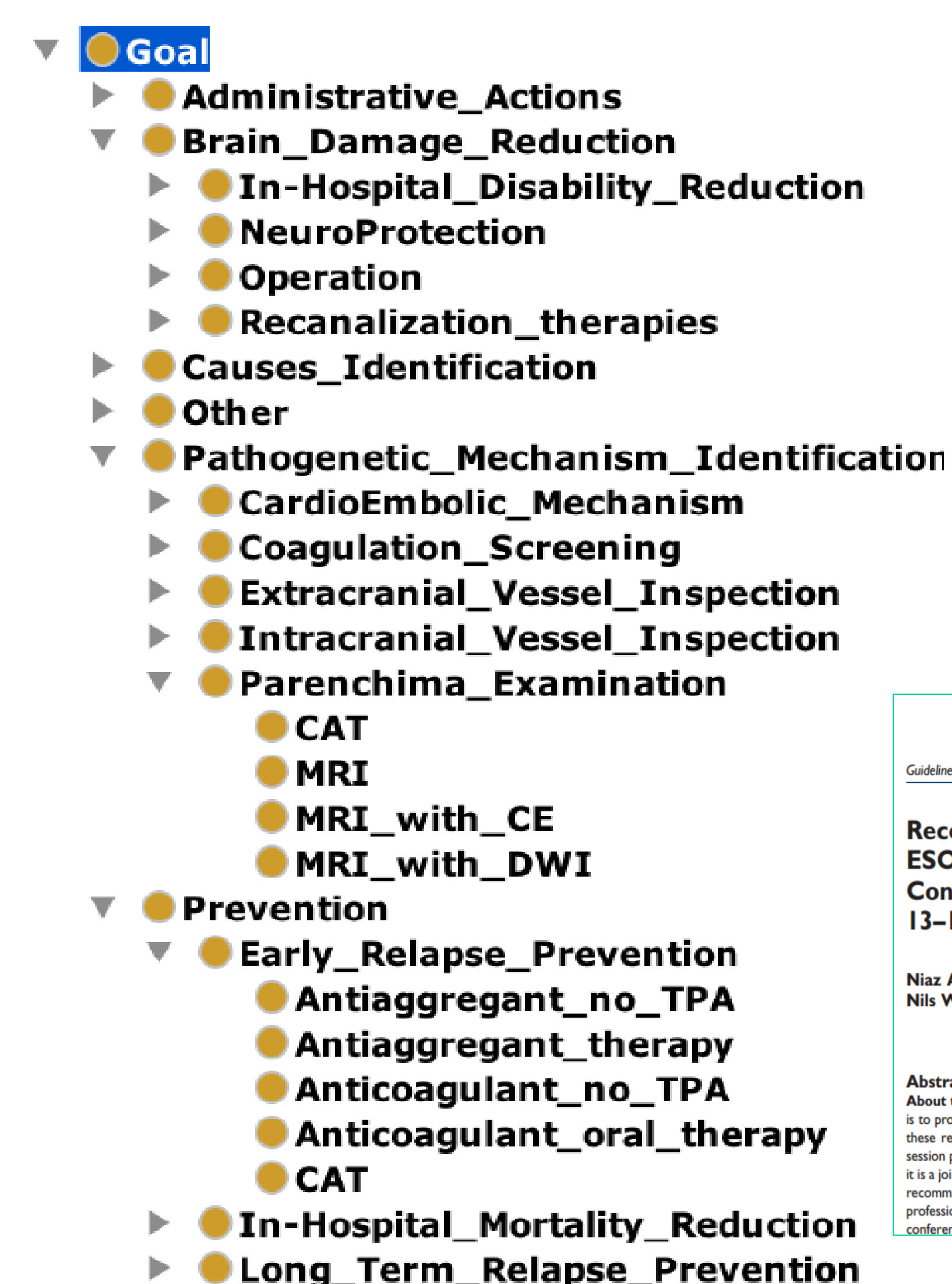
- Many organizations record information about the executed business process instances in the form of an **event log**
- Sequence of actions actually executed (**process traces**)
- Source of information for BPM tasks
 - **Trace comparison and retrieval**
 - Operational Support
 - Derive indications on running trace completion time/costs/resources
 - Suggest next actions or support routing decisions
 - Process mining
 - More readable process model if learned on homogeneous traces
- CB Retrieval step can be adopted in this context

2. Abstracted trace semantic comparison

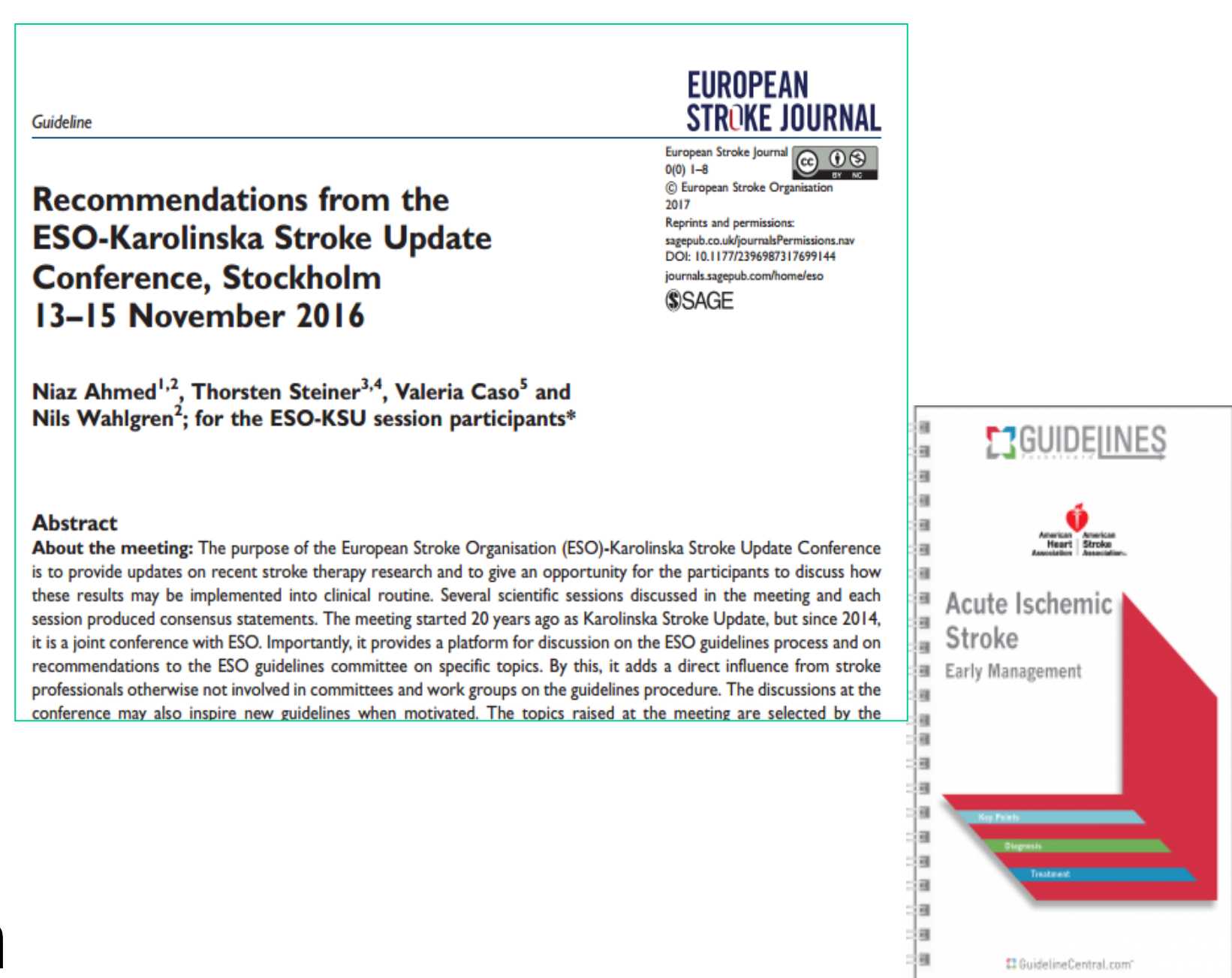
- Current approaches adopt a purely syntactical analysis, and compare trace actions referring to their names
- A semantic approach can provide more reliable results, by taking into account conceptual relatedness
- We propose a framework for
 - **Abstracting** trace actions on the basis of their **semantics**
 - **Comparing** abstracted traces exploiting **domain knowledge**

3. Knowledge-based abstraction

- By mapping trace actions to leaves of a taxonomy, we can abstract them up to the desired level
- We can obtain an abstracted trace as a sequence of **macro-actions**, merging subsequent actions that abstract as the same ancestor
 - Managing **delays**
 - Considering **interleaving actions**

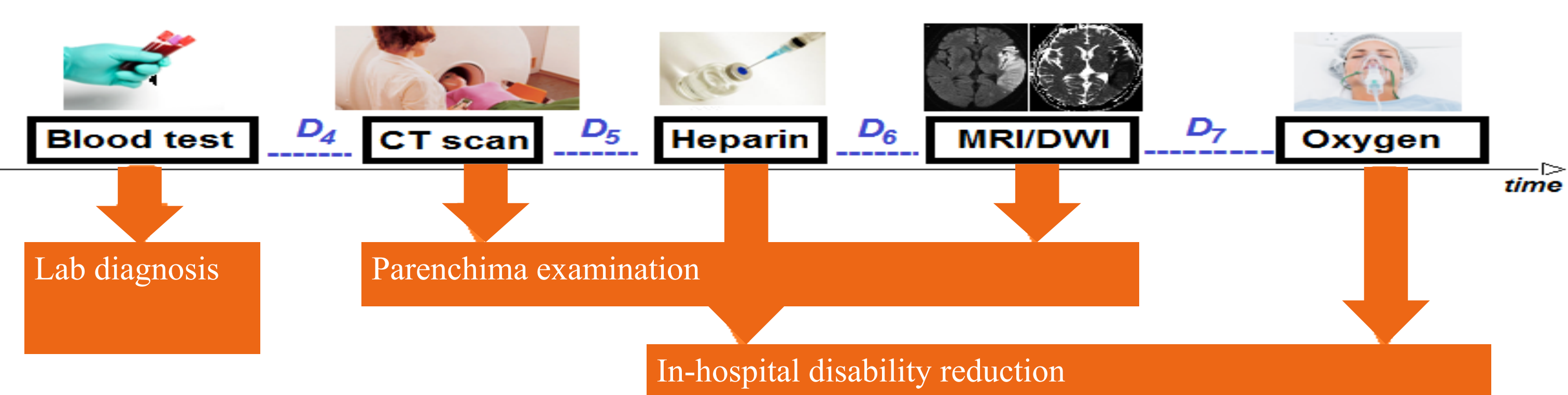


Stroke goal/subgoal taxonomy



4. Abstraction mechanism

- User-selected abstraction level
- $(D_5 + D_6) < \text{Delay threshold}$
- $|\text{Heparin}| < \text{Interleaving threshold}$



5. Trace comparison

- Sequence of macro-actions
 - Semantic edit distance
 - Optimal alignment

$$\text{CAT-MRI: } \frac{(1+1)}{(1+1+2*2)} = 0.3 \quad \text{MRI-Antiaggregant_ther: } \frac{(3+3)}{(3+3+2*0)} = 1$$

$$dt(\alpha, \beta) = \frac{N_1 + N_2}{N_1 + N_2 + 2 * N_3}$$

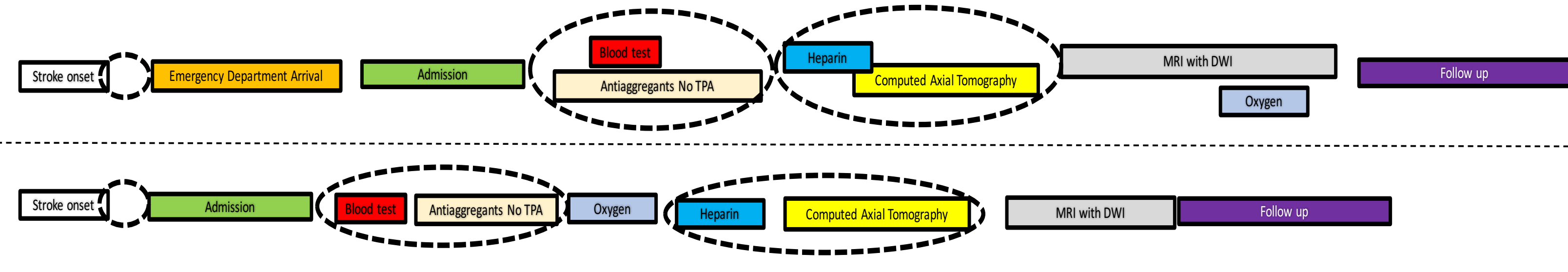
- Temporal information
 - Aligned macro-actions duration
 - Temporal constraints between pairs of aligned macro-actions

$$\text{interval}_d(A, B) = \frac{|\text{len}_A - \text{len}_B|}{\text{maxlen}}$$

$$\text{ngraph}_d(i, j) = \frac{\text{path}(i, j, G)}{\max_{k, l \in G} \{(\text{path}(k, l, G))\}}$$

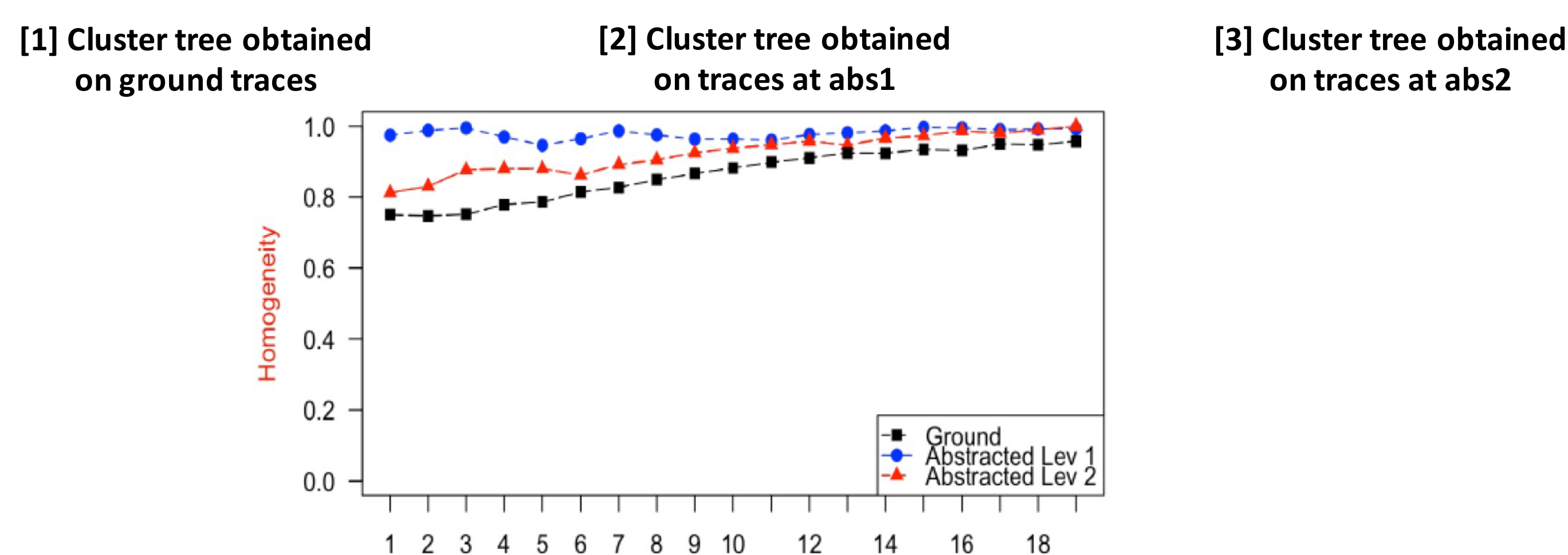
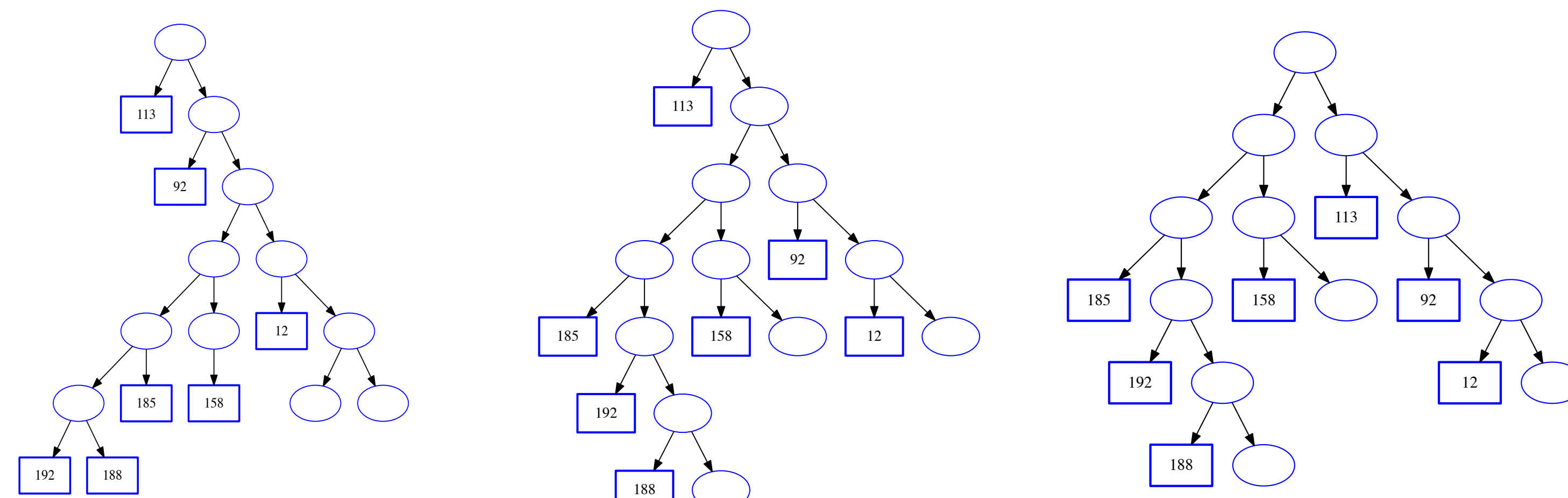
- Abstraction penalty

$$\text{delay}_p(A, B) = \frac{|\text{delay}_A - \text{delay}_B|}{\text{maxdelay}} \quad \text{inter}_L_p(A, B) = \frac{|\text{inter}_A - \text{inter}_B|}{\text{maxinter}}$$



6. Experiment

- Clustering on stroke management traces
 - Unweighted Pair Group Method with Arithmetic Mean
- Hypothesis
 - «closer» inter-cluster items (higher homogeneity): more robust clustering results
 - Preserved capability of isolating outliers



7. Future work

- Abstract trace comparison can be adopted to retrieve similar traces in operational support
- Abstract trace clustering can be exploited as a pre-processing step in process mining
 - Indeed, ground medical processes (learned on trace at the same level of taxonomy leaves) are typically «spaghetti-like»: they presents an extremely large number of nodes and edges which make it hard to analyse the model and identify details
- Methodological work:
 - Enrich knowledge representation: towards an ontology
 - Connected to SNOMED terms
 - Considering viewpoints other than goals
 - Introduce additional penalty definitions
 - Develop a rule-based approach to introduce «context» in the abstraction mechanism