VISUALIZATION METHOD EFFECTIVENESS IN ONTOLOGY-BASED INFORMATION RETRIEVAL TASKS INVOLVING ENTITY EVOLUTION

AKRIVI KATIFORI, COSTAS VASSILAKIS, GEORGE LEPOURAS, ELENA TOROU, CONSTANTIN HALATSIS

PRESENTER: IOANNA LYKOURENTZOU

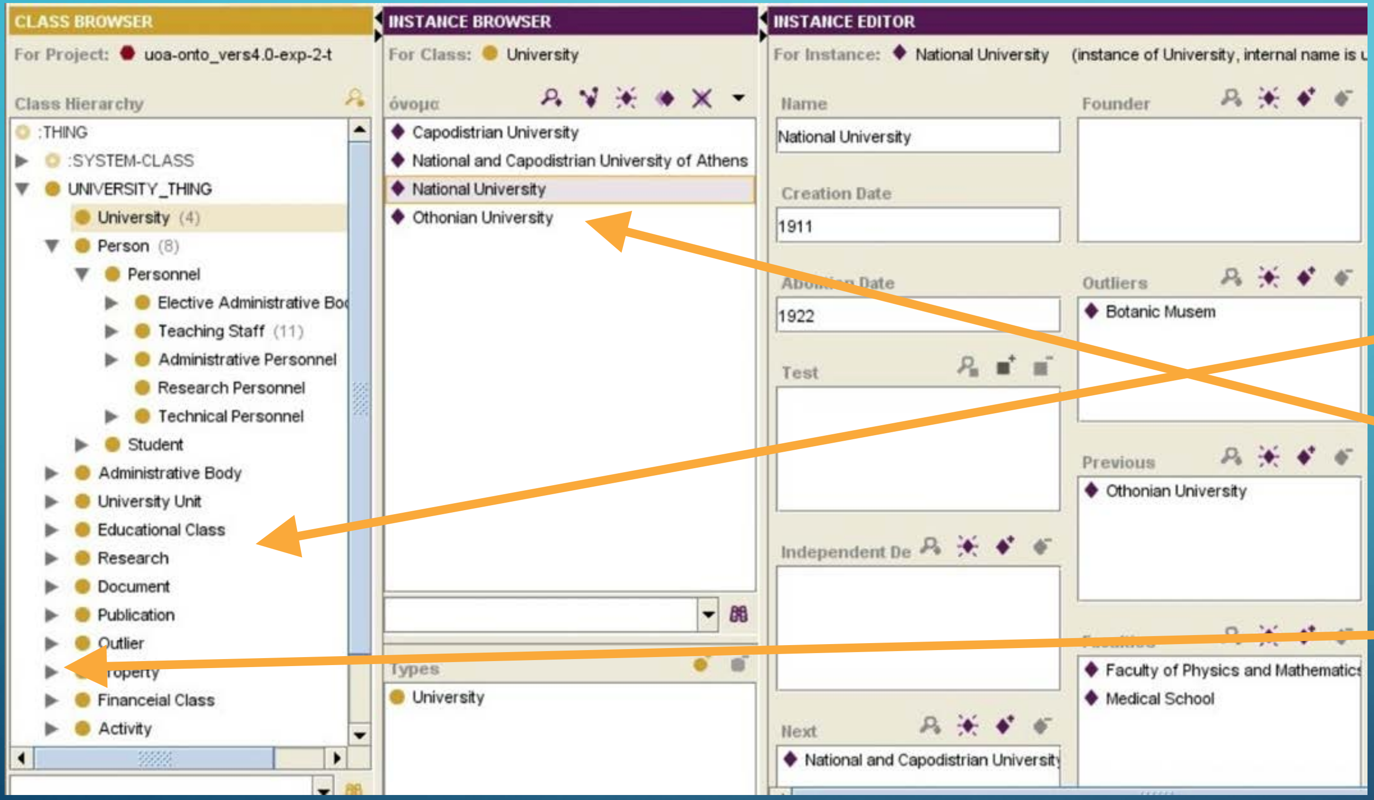
THE PROBLEM

- We have too much content / data!
  - Management, search and retrieval are becoming increasingly difficult
  - This is particularly true in big repositories where historical/time-evolving data is stored, e.g. historical archives
- Semantic data organization and indexing, coupled with appropriate visualizations, can offer to domain experts efficient access to high quality information
- Ontologies can be very useful to this end
  - They can be used for semantic data organization, and also handle time-evolving data
  - Ontology visualization methods can allow users to see, explore, and understand large amounts of information at once
  - These features can facilitate the information retrieval process

WHAT WE HAVE DONE

- A user study focusing on the effectiveness of current ontology browsing and visualization methods in the following context:
    1. when the information model, and therefore the ontology, involves entity evolution, i.e. the same entity may appear in various versions (e.g. a member of the academic community may initially appear as a student, then as a PhD candidate and subsequently as a lecturer)
    2. when the ontology visualization methods are used for supporting users in information retrieval tasks
- We have explored four well-established ontology visualization paradigms: indented list, zoomable, focus+context and node-link+tree.
- Effectiveness and sufficiency of visualization methods are assessed through both objective and subjective data

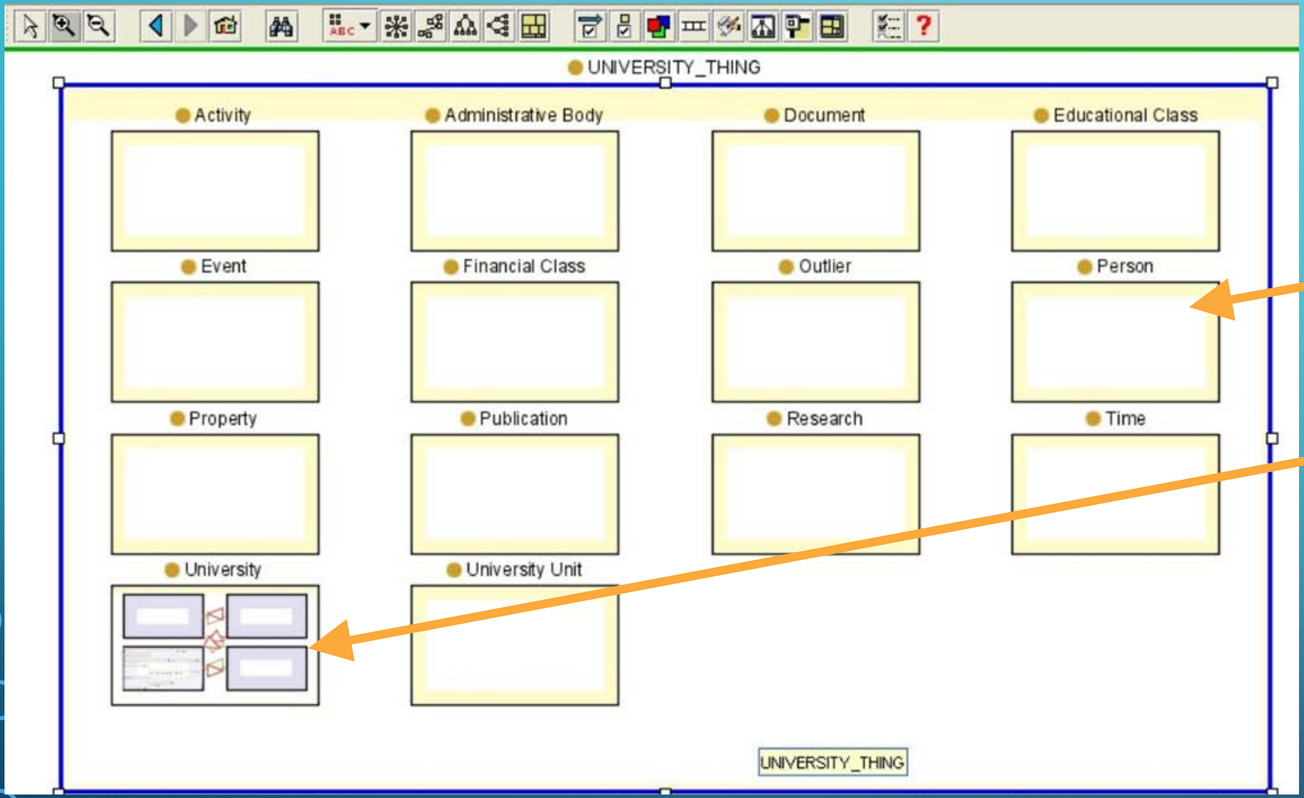
EXPERIMENT DESCRIPTION: VISUALIZATION METHODS EXAMINED (1/4)



Class Browser

- Representative of the "indented list" visualization approach
- The is-a taxonomy is represented as an indented list
- Instances of the selected class are displayed on a separate pane
- Class hierarchies can be expanded or retracted through appropriate handles
- Familiar method, used widely in file browsers

EXPERIMENT DESCRIPTION: VISUALIZATION METHODS EXAMINED (2/4)

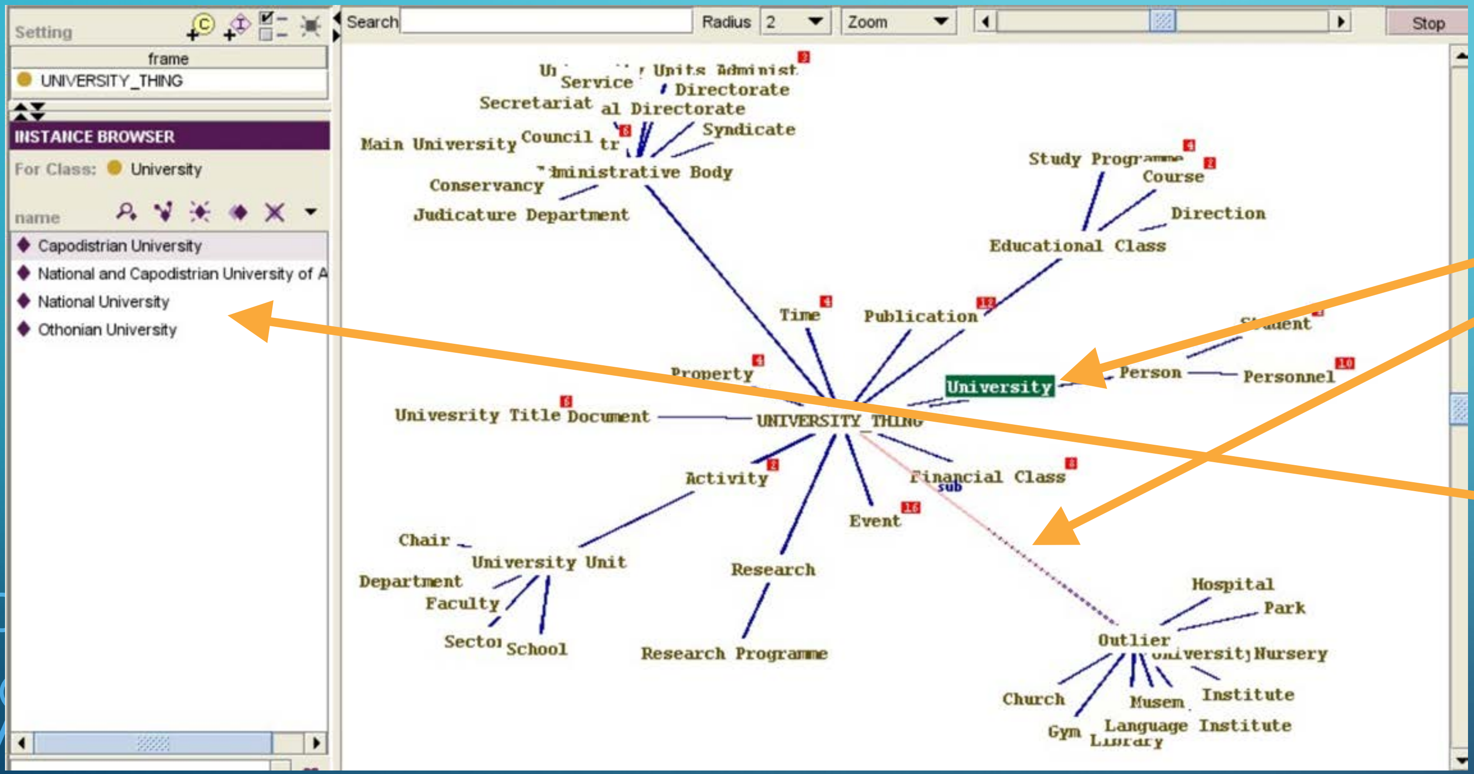


Jambalaya

- Representative of the zoomable view visualization paradigm
- When a class is expanded, the current view is replaced by one depicting its subclasses
- Instances are represented as nested nodes in their corresponding class in the graph
- Role relationships can also be presented

- Also familiar from its use in file browsers

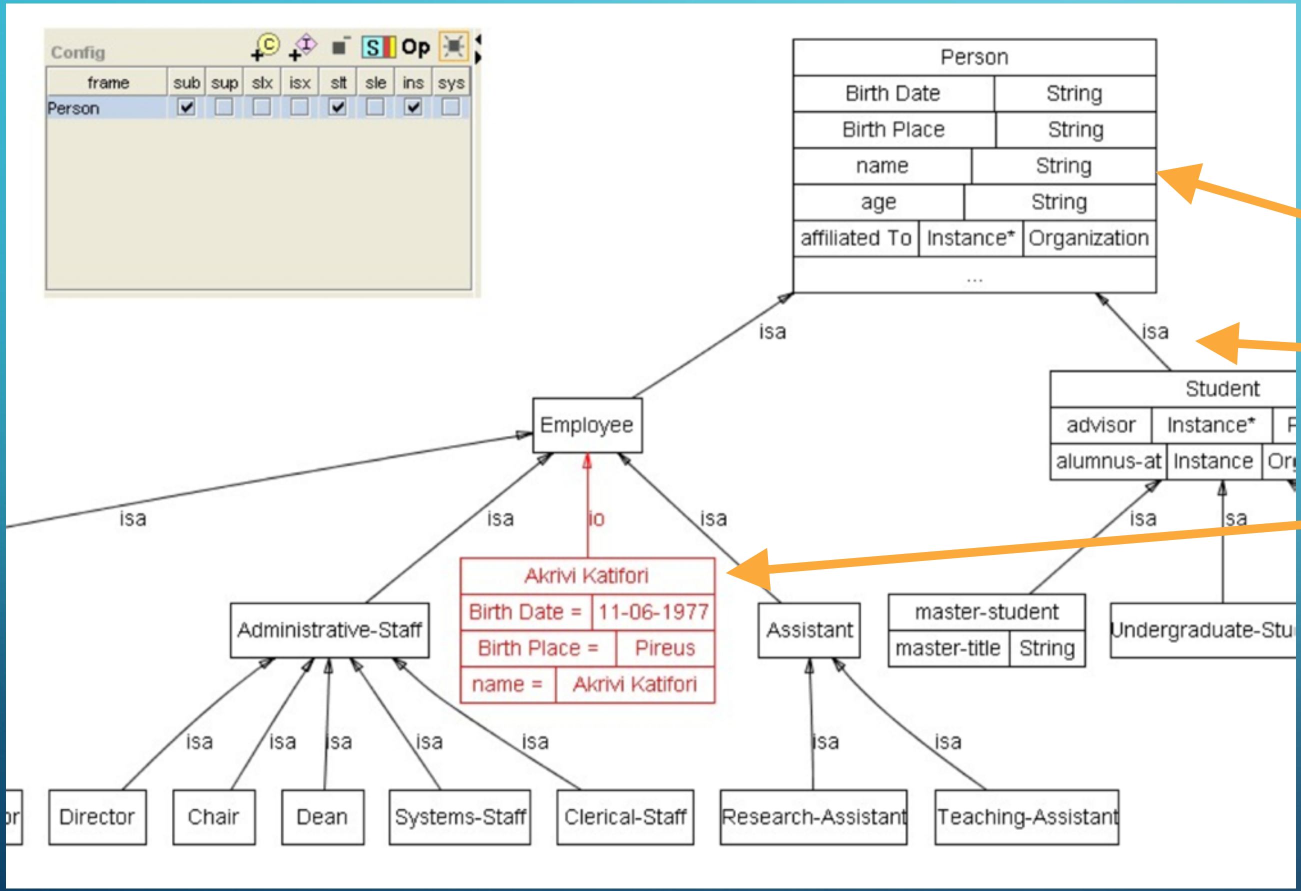
EXPERIMENT DESCRIPTION: VISUALIZATION METHODS EXAMINED (3/4)



TGViz

- Representative of the focus+context visualization paradigm
- Classes are mapped to nodes and is-a relationships to edges
- Classes can be expanded or retracted or even hidden
- Instances of the selected class are displayed in a separate pane
- When a node is moved, nodes linked to it also move along with it
-

EXPERIMENT DESCRIPTION: VISUALIZATION METHODS EXAMINED (4/4)

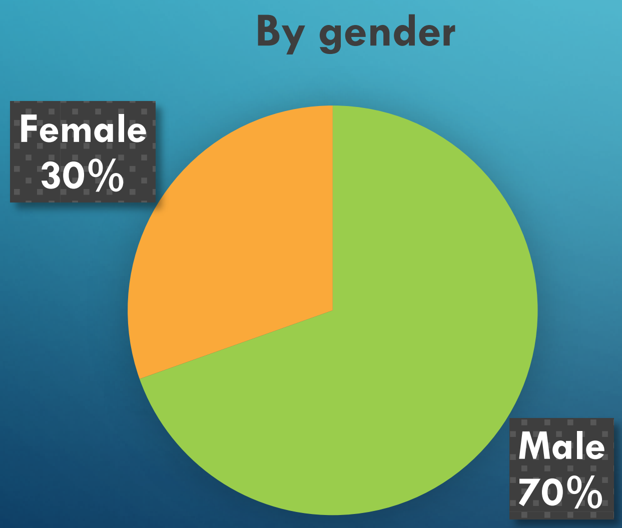


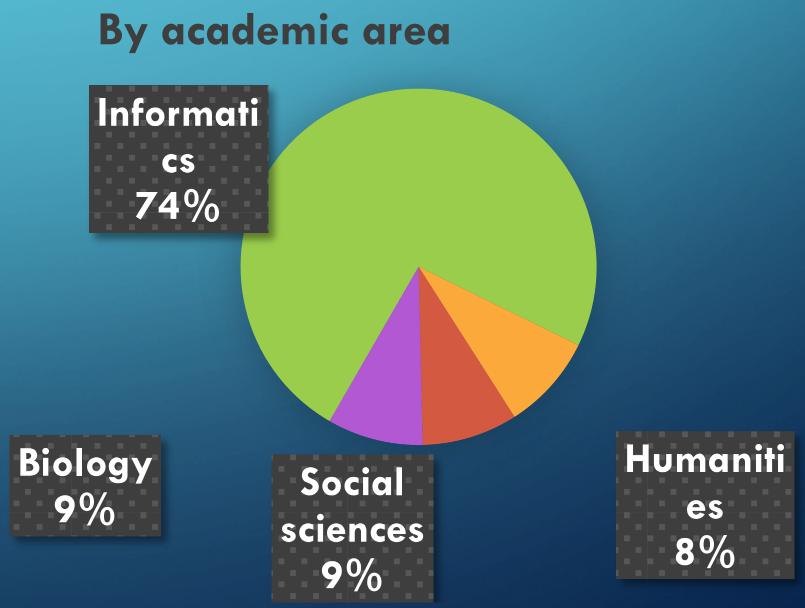
Ontoviz

- Representative of the node-link+tree visualization paradigm
- Classes are displayed as tree nodes, with the tree hierarchy corresponding to the is-a relationship taxonomy
- Instances are displayed in the same pane using a different color
- The information presented in the graph (classes, instances, slots etc.) can be customized via the configuration panel

EXPERIMENT DESCRIPTION: EVALUATION USER GROUP

- 23 participants in total
  - All participants had at least basic computer expertise
  - None of the participants claimed to be an experienced ontology designer
    - Eight claimed to understand the notion of “ontology” to some extent





EXPERIMENT DESCRIPTION: THE ONTOLOGY USED

- The University of Athens Historical Archive ontology
  - Stores the current state of the University along with past states, containing thus *temporal information*
  - Temporal information can be incomplete (i.e. not all information is recorded) or uncertain (i.e. it may be correct with some probability; contradictory pieces of information can be also found)

Ontology statistics	
# classes:	205, sparsely populated; 2/3 of them have no direct instances
# instances:	599
Maximum depth of is-a taxonomy:	5 classes
Mean depth of is-a taxonomy:	3 classes
# distinct slots:	176, with more than half of them describing relations between classes

EXPERIMENT DESCRIPTION: INFORMATION RETRIEVAL TASKS

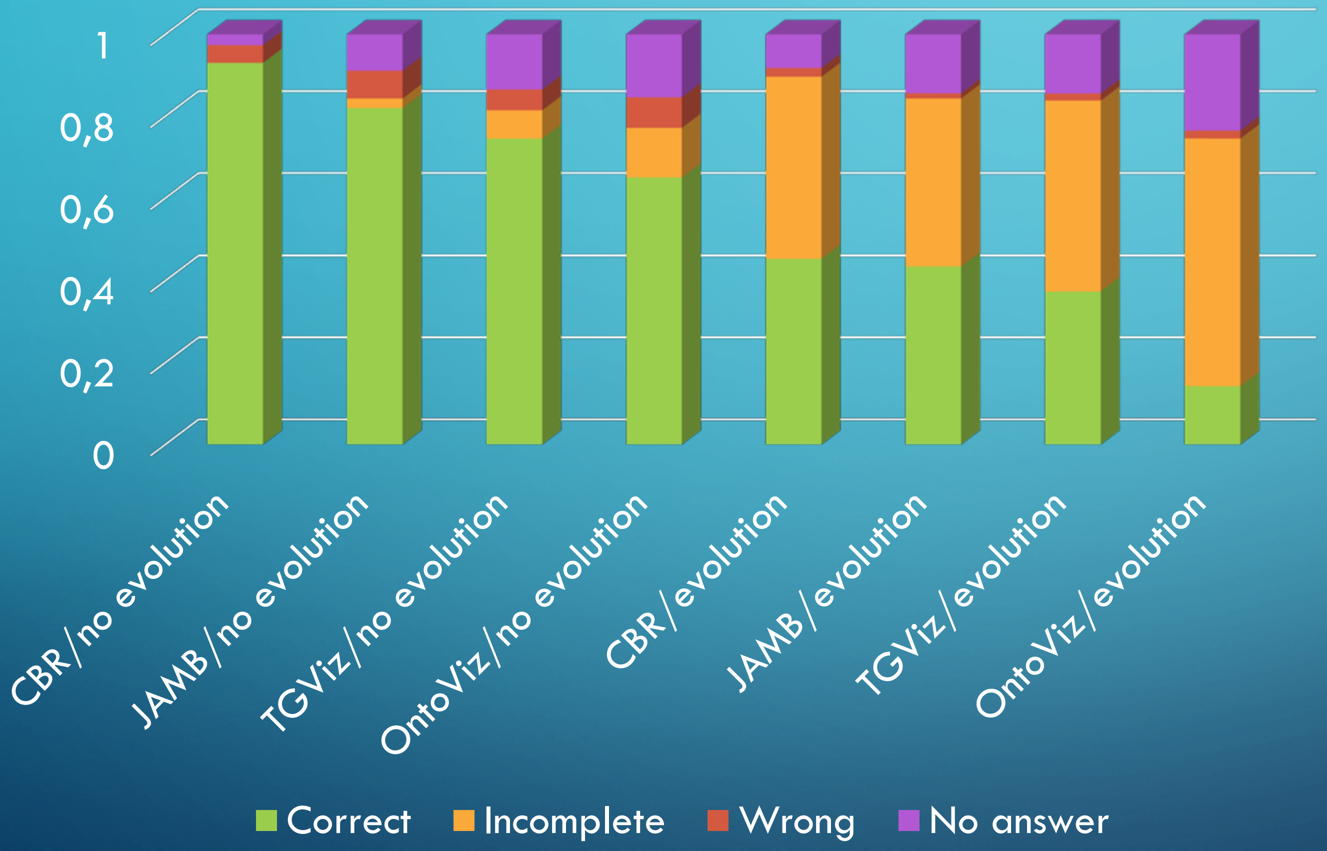
- To obtain a representative set of IR tasks, the list of queries posed to the Historical Archive was analyzed
- Also, typical IR tasks performed through visualizations were sourced from the bibliography

Insight gained
58% of queries do not involve entity evolution
42% of queries involve entity evolution (24% person biographies and 18% historical evolution of institutions/organizations)

- Set of IR tasks used in the experiment:

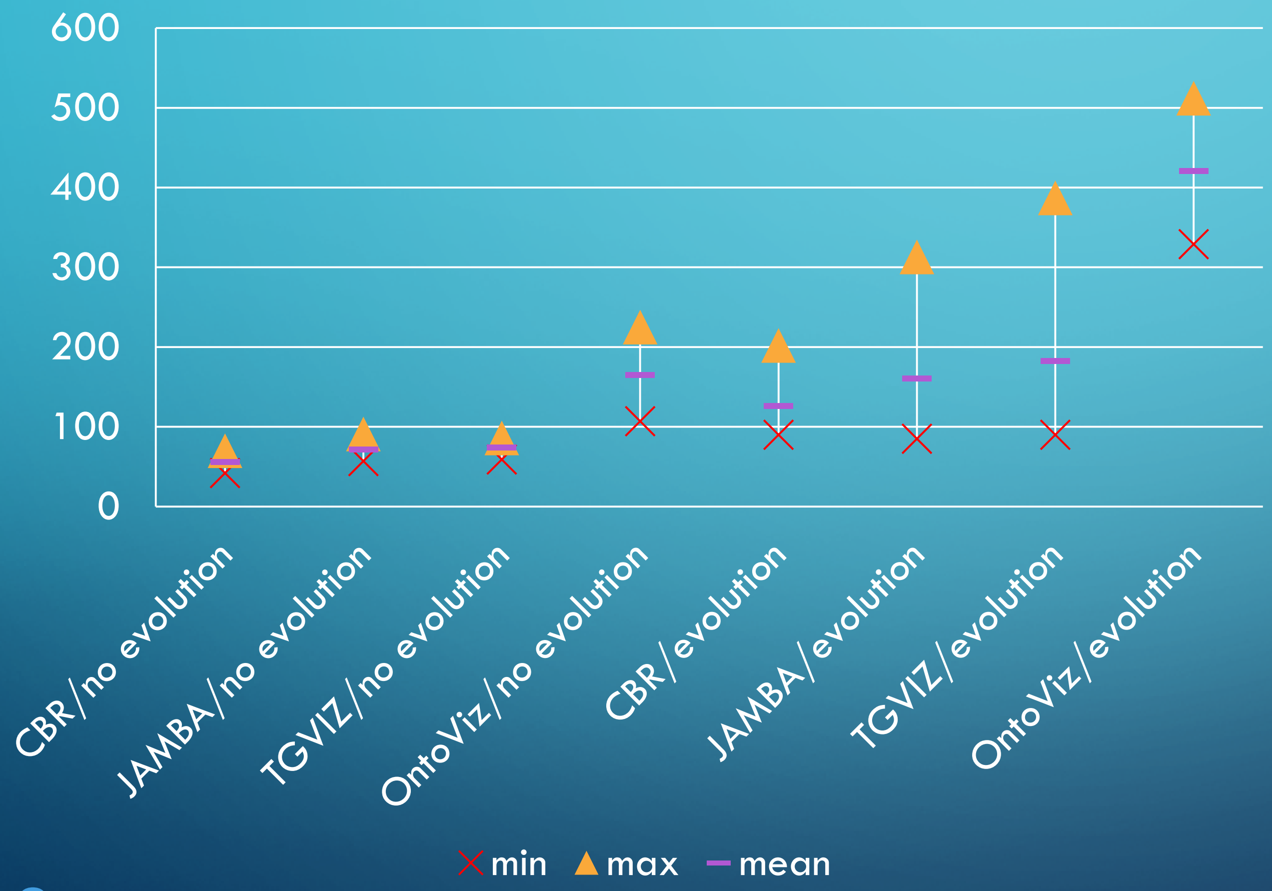
Insight gained
3 tasks not involving evolution; mainly locate a class or an instance and possibly traverse is-a or role relationships
4 tasks involving evolution; the participant has to find multiple (or all) versions of the same entity in the ontology, and possibly perform some calculations over their data

EXPERIMENT RESULTS: CORRECT, INCOMPLETE, WRONG AND NO ANSWER PERCENTAGES



- Satisfactory performance is observed in queries not involving evolution, with class browser having a performance lead
- Performance drops dramatically when entity evolution is involved.
  - Performance is especially low when calculations need to be performed on data values extracted from ontology versions retrieved, e.g. compare dates
  - A lot of answers were incomplete, since users failed to locate the requested information

EXPERIMENT RESULTS: TIME TAKEN FOR IR TASK COMPLETION



- Tasks involving evolution may need up to 3x more time
- Class browser has again a performance lead, with OntoViz being the worst performer
  - Its poor performance though may be owing to deficiencies of the particular tool and not be representative of the visualization approach
- Findings provide an indication that IR tasks involving evolution are not adequately supported by current visualizations and novel ones are needed.

DISCUSSION - CONCLUSIONS

- Current ontology visualization methods appear to handle non evolution-related IR tasks satisfactorily, however evolution-related tasks are not adequately supported
- Insight gained from the experiment shows that the performance lag is owing to various reasons including:
  - Lack of visualization ability to present entity evolution timelines
  - Improper modelling entity evolution in the ontology
  - Inability of the tool to support the user's browsing pattern (e.g. some users prefer systematic browsing, others random/quick browsing; some users avoid unknown territories; and so forth)
  - Intricacies of the visualization tool; for instance OntoViz was ranked as the less user-friendly tool due to its lack if interactivity
  - Lack of a complementary query mechanism to alleviate the burden of performing caclulations

FUTURE WORK

- Perform more experiments aiming to study the effect of each of the identified impeding factors
- Experiment with specific tool features such as animation
- Study the appropriateness of combining different visualizations to achieve better performance

THANK YOU!

Comments?

Questions?