

Visual Ontology Alignment System – An Evaluation

V. Sabol¹, W. O. Kow³, M. Rauch¹, E. Ulbrich¹, C. Seifert², M. Granitzer², and D. Lukose³

¹Know-Center, Inffeldgasse 13/VI, 8010 Graz, Austria

²Faculty of Computer Science and Mathematics, Passau University, Innstrasse 33, 94032 Passau, Germany

³Artificial Intelligence Centre, MIMOS Berhad, Technology Park Malaysia, 57000 Kuala Lumpur, Malaysia

Abstract

Ontology alignment is the process of mapping related concepts from different ontologies. A lot of research effort has been invested in development of algorithmic methods supporting automatic discovery of mappings between ontological concepts. However, automatic alignment remains potentially prone to errors especially with large real-world ontologies, demanding intervention of domain experts. We therefore created a semi-automatic tool including algorithmic alignment methods and an interactive visual interface. Visualisation components included in the interface support experts in navigating the concept space and reviewing the automatically generated mapping suggestions. An experiment with 15 test users was performed to evaluate whether, and in which cases the use of visualisation is beneficial compared to a user interface employing standard GUI widgets. The results indicate that users typically executed tasks slightly faster with an interface using standard widgets, but an interface which includes a visualisation component providing overview, filter and narrowing-down functionality achieved higher rates of successful task completion.

Categories and Subject Descriptors (according to ACM CCS): H.5 [Information Interfaces and Presentation]: User Interfaces—User-centered design

1. Introduction

Since the advent of semantic technologies, more and more organisations and people share their knowledge in the form of ontologies. Interoperability issues may arise when ontologies express the same knowledge in different ways, for example due to the use of different terminology, divergent points of views, or differing levels of model granularity. Ontology alignment is the process of finding mappings between related concepts from two or more different ontologies. Incompatibilities between ontologies and the need for interoperability between the rapidly growing number of systems using semantic technologies lead to an increased need for ontology alignment.

There have been a variety of algorithmic solutions proposed for automatic alignment based on different approaches, beginning from simple string matching, over linguistic methods, reasoning, machine learning techniques, to specialised techniques such as similarity flooding. For more details on various algorithmic methods for ontology alignment see [ELBB*04], [ES07] and [GS08]. To compare the performance of different methods, the Ontology Alignment Evaluation Initiative [OAE11] organises a yearly event fo-

cus on evaluation of ontology alignment algorithms and systems. By following the improvement gains on a yearly basis it can be observed that, despite an increasing research effort being invested into new, more complex algorithmic methods, the gains are tangible but diminishing.

Automatic ontology alignment methods may be a viable solution in some cases, but they are likely to underperform in scenarios where deep understanding of a specific application domain is required, large ontologies are aligned and high quality of the computed mappings is expected [Rah11]. In order to overcome these challenges, human experts should be adequately integrated in the alignment process [KL08] to make use of their wide knowledge and rich experience. Semi-automatic ontology alignment systems provide the possibility to improve on the automatic alignment techniques through the involvement of domain experts, who can decide whether an automatically computed mapping is correct or not.

However, domain experts are rarely knowledgeable in semantic technologies and thus require easy to use tools to perform this kind of tasks. Visual methods provide means for exploration and analysis of large amounts of complex infor-

mation by making use of the powerful human visual capabilities [TC05]. Use of visualisation to support ontology alignment has been discussed and advocated by several authors, such as in [LSR*08]. In [GSK*10] a survey of visual ontology alignment systems is given and compared to a list of requirements collected from various user studies, concluding that no system currently fulfils all requirements.

In this paper we present the "Semantic Mediation Tool" (SMT), a semi-automatic visual ontology alignment system using algorithmic methods to compute an initial set of mapping suggestions, which are then reviewed by experts using a visual user interface. The focus of the presented work is two-fold: i) design of the user interface with the goal of fulfilling all user requirements, and ii) evaluation of performance of visualisation for semi-automatic ontology alignment tasks.

This paper is structured into eight sections. Section 2 gives a short overview of visual ontology alignment tools, Section 3 provides an architectural overview of our system, and Section 4 outlines the algorithms we used for ontology alignment. Section 5 introduces the list of user requirements and presents the visual interface of our tool. Section 6 describes the investigation procedure we used to find out whether use of visualisation is suitable for performing selected tasks. Evaluation results are reported and discussed in Section 7. Finally, Section 8 sums up the results of our evaluation and discusses future work.

2. Related Work

In visually supported, semi-automatic ontology alignment systems the involvement of humans necessitates an easy to use user interface, where visual components are used to support overview, pattern recognition and navigation functionality. In our previous work [GSK*10] we compiled a survey of semi-automatic, visually supported ontology alignment systems. The majority of available systems can be subdivided into three main categories depending on the visual paradigm employed by the user interface: i) interfaces based on linked tree widgets, ii) interfaces based on graph visualisation, and iii) treemap-based interfaces. In the following we give an brief overview of existing visual ontology alignment systems grouped into these three categories.

Interfaces based on linked trees use the standard tree widget to present the class hierarchy of the two ontologies, with both trees being placed next to each other. Mappings are shown as lines or curves connecting concepts in different trees, such as the AgreementMaker [CSMB07], COMA++ [ADMR05] and COGZ [FS07, FBG09]. The interface of PROMPT [NM03] is similar, however it shows mappings in a table instead of connecting the trees.

Since ontologies are graph structures, graph-based visualisations are a natural fit for visualising ontologies. Tools using graph visualisation methods to represent ontology nodes and generated mappings include Optima [KD08],

HOMER [UGM07], PROMPT [NM03], AIViz [LS06], which provides a combination of trees and graph views, and [dSDdMR06] which uses hyperbolic geometry to reduce clutter.

The treemap [Shn91] is a visual representation providing an overview of hierarchical structures, where nodes are represented as nested rectangles. The size and colour of each rectangle encode properties of the corresponding class, such as the number of leaves and the amount of found mappings. COGZ [FS07, FBG09] includes treemaps to provide an overview of the class taxonomy and uses colours to show which regions contain many or few mappings.

Our survey [GSK*10] summarises the findings of user studies, such as [FNS06, FNS07], into a list of requirements for interactive ontology alignment tools (see Section 5 for details). Subsequently, these requirements were compared to the available visual ontology alignment systems and it was concluded that no current system fulfils all requirements. As a consequence we proposed a visual interface which was designed by following the recommendations and requirements compiled in the survey.

In [KSG*11] a very brief overview of an early version of the resulting "Semantic Mediation Tool" (SMT) was presented, but no details were disclosed and no evaluation results were presented. SMT uses an information landscape visualisation to provide an overview of concepts from both ontologies and employs graph visualisation components for showing details. The springScape system [EBJ06] uses an information landscape for visualizing multiple data sources (ontologies) in the context of microarray and contextual bioinformatic data. However, springScape does not address ontology alignment. Also, in contrast to SMT, the employed layout algorithm does not scale well and may produce different results over subsequent runs for the same data set (reproducibility problem). The remainder of this paper gives a detailed description of our system and provides results of the performed user evaluation.

3. Architecture

SMT is conceived as a server-client system implemented in Java. Alignment algorithms are executed on the server and are implemented as so-called Matcher components. Currently two matcher implementations exist, a linguistic and a statistical Matcher (see Section 4 for algorithm details), but due to a standardised Matcher API additional alignment algorithms can be easily added to the system. Matchers have access to ontologies to be aligned through an API which encapsulates triple stores (currently Jena [CDD*02] and AllgroGraph [W3C09] engines are supported). When a pair of ontologies is aligned, the list of concept mappings computed by a Matcher is stored under a user-specified name in the Mediation Repository and can be retrieved and reviewed by the user at a later time. Additionally, a full-text search index and visualisation geometry data (see Subsection 4.2) are

generated and stored in the Mediation Repository together with the list of concept mappings.

The client implements a visual user interface which connects to the server and displays the data stored in the Mediation Repository. The user can review the computed concept mappings and use the visualisations to explore the concept space and the generated mappings. The user interface consists of the following main components:

- Mapping table: a table component for displaying the suggested mappings.
- Ontology browsers: two graph visualisation components for exploring concept properties and navigating in the ontologies.
- Information landscape: a visualisation providing an overview of the entire concept space, supporting narrowing down to regions of interest.
- Ontology trees: two tree components displaying the class hierarchies and concepts of the aligned ontology pair.

A coordinated multiple view framework works behind the scenes to synchronise the different views and ensure that user actions, such as selection or filtering, performed in one component are adequately reflected in other components.

4. Algorithms

This section describes the two ontology alignment algorithms used in our system, a linguistic method and an unsupervised learning-based method. The algorithms employed in the second method are also being used for computing the geometry for the information landscape visualisation. It should be noted that, since the focus of this paper is on visualisation, an into-depth description and evaluation of the algorithms is not included.

4.1. Linguistic Method

This method uses an external taxonomy to measure the semantic distance between the two concepts based on the name of the concept. In this case, we are using the WordNet [Uni10] taxonomy, but the algorithm can be adapted to use any taxonomy. The Wu-Palmer measure [WP94] is used for calculating the similarity values of the mappings. Using this measure, similar concepts (members of the same WordNet synset) would have a score of 1. Concepts that share the same parent will have a score slightly below 1, while concepts that share the same grandparent will have a lower similarity. The distance between the common ancestor and the root of the taxonomy is also taken into account, so that siblings higher up in the tree (more general) will have a lower similarity compared to the more specific ones near the bottom of the tree.

4.2. Unsupervised Learning-based Method

The second alignment method is a machine learning-based approach utilising algorithms implemented in the

Know-Center’s KnowMiner knowledge discovery framework [KSM*09]. The algorithm consists of three steps: concept vectorisation, concept clustering, and mapping finding. In the first step every concept is converted into one or more feature vectors using following information: i) concept label, ii) concept description (if available), iii) neighbouring concept labels, and iv) relationships connecting to neighbours. Additionally, WordNet is used to extend the label information with synonyms (as well as hypernyms and hyponyms, if desired). In this way up to four different vector spaces are spawned which are used to compute the cosine similarities between concept pairs. It should be noted that, besides concept labels and descriptions, structural information (i.e. relationships and neighbours) and linguistic information (i.e. label synonyms) also contribute to the similarity. A compound similarity value over all spaces allows for adjusting the weight of each space. Currently the weights are fixed, but in a future version of the algorithm they could be automatically adjusted depending on the specifics of the ontologies to be aligned.

Once concept vectors are available, the concepts from both ontologies are clustered using a scalable hierarchical clustering algorithm [MSG10] running in $O(N * \log(N))$ time and space, N being the total number of concepts. The algorithm creates a balanced hierarchy of clusters by recursively applying a modified x-means [PM00] clustering method. As a result, similar concepts will be gathered in the same clusters, even when they originate from different ontologies. To avoid comparing all concept pairs when finding mappings, which would lead to a quadratic execution time, mappings are found by inspecting only pairs of concepts assigned to the same cluster (or sub-cluster). By choosing sub-clusters deeper in the hierarchy in such a way that the number of concepts within the branch is smaller than a fixed threshold $C \ll N$, the number of comparisons performed for each concept is $O(C)$ (i.e. constant) resulting in the running time of the mapping-finding step being linear with the total number of concepts N .

The hierarchy produced by the clustering algorithm is also used for computing the similarity layout (i.e. a layout where similar objects are placed spatially close to each other) and the geometry needed for the information landscape visualisation. The hierarchy computation of the 2D similarity layout and cluster area subdivision are performed by a scalable projection algorithm having the same $O(N * \log(N))$ time and space complexity as the clusterer [MSG10]. The algorithm proceeds recursively along the cluster hierarchy: First, the top-level clusters are positioned using a force-directed placement method [FR91] producing a layout where similar clusters are placed close to each other. Then, each cluster is assigned a polygonal area using Voronoi area subdivision [Aur91]. The method proceeds recursively by positioning sub-clusters within their parent cluster’s area and then assigning Voronoi areas to the sub-clusters. Recursion stops when the bottom of the hierarchy, i.e. the concepts, are

reached and positioned. Scalability of the algorithm is due to the fact that at each hierarchy level the number of direct children of a cluster has a strict upper limit, which is guaranteed by the clustering algorithm. As a result, the force-directed placement method, which is known to compute good similarity layouts but scales poorly, is only applied on small data chunks (consisting of, for example, 20 elements) on which it operates very quickly.

5. Visual User Interface

As already mentioned in Section 2, the design of our system is based on the summary of user requirements for interactive alignment tools provided in [GSK*10]. The requirements are as follows:

- 1) Presentation of automatically generated mapping suggestions including an estimated confidence for each mapping.
- 2) Exploration and navigation of ontologies providing details on every concept.
- 3) Overview of the concept space and the alignment results.
- 4) Capability to narrow down to the area of interest.
- 5) Filtering based on features of concepts and mappings.
- 6) Confirming, rejecting and editing of automatically generated mappings.
- 7) Collaboration via commenting, tagging, annotating etc.
- 8) Ability to partition the reviewing task into chunks assignable to team members.
- 9) Saving and loading of user's changes.

Visual techniques appear particularly useful for addressing Requirements 2, 3 and 4, while Requirements 5, 6 and 8 can also benefit from visualisation methods. Therefore we designed a visual user interface, shown in Fig. 1, providing support for the above requirements. The two aligned ontologies in this example are the **mouse** anatomy ontology (red), and the **human** anatomy ontology (green), both provided by the Ontology Alignment Initiative [OAE11]. The colour used for each ontology is configurable to support users with red-green colour blindness. The user interface consists of the following main components: a table of mappings (up-left) addressing Requirements 1, 6 and 7, two ontology browsers (on right) addressing Requirement 2 and supporting Requirement 6, and an information landscape (in the centre) addressing Requirements 3 and 4 and supporting Requirement 5 and 8. Requirement 5 is fully supported by a search facility (upper-right corner). Requirement 8 is supported by the capability to select a subset of the mappings, typically using the information landscape (see Subsection 5.2), and assigning it as a task to an (expert-)user (second button from the right in the tool bar), who will be able to review and modify only the assigned mappings. Monitoring of task progress is also possible (rightmost button in the tool bar). Finally, saving and loading of user changes (button in the upper left corner of the tool bar) is also provided satisfying Requirement 9. Therefore, in contrast to other visual ontology alignment systems (see Section 2), the proposed user interface, which

is described in the rest of this section, fulfils all requirements listed above.

Using an information landscape visualisation in the context of ontology alignment is a novel concept, which needs to be evaluated (see Section 6). In order to compare the information landscape with the standard tree widgets which are typically used in comparable tools, we included the possibility of displaying and navigating ontology class hierarchies using a pair of trees (shown in Fig. 2). These ontology trees address the same requirements as the information landscape.

5.1. Mapping Table

The mapping table (top-left in Fig. 1) displays mappings discovered by the alignment algorithms. Table columns show the name of the mapped concepts (with coloured icons encoding ontology membership), a similarity score between 0 and 1, a status column (suggested, accepted, rejected), and a reviewer column (initially empty). By default, mappings are sorted by their estimated similarity (confidence), but sorting by any other column is possible. Initially all mappings are in suggested state. The user can review the mappings, change their status to accepted or rejected (using buttons in the tool bar), add a comment, and save the performed changes. Navigation in the table takes place by paging in groups of 10, 20 or 50. Filtering of concepts and mappings is possible depending on the mapping state (drop-down list in the tool bar) and by navigation and selection in the information landscape (see next subsection). Concepts and mappings not fulfilling the filter criterion are removed from the mapping table. Additionally, backed by a built-in retrieval subsystem, our tool supports searching as well as fast highlighting and filtering based on full-text and Boolean queries. To optimise the reviewing, various keyboard shortcuts are provided together with a single-click function for accepting the selected concept pair and rejecting all other mappings containing one of the concepts.

5.2. Information Landscape

In the centre of Fig. 1 the information landscape visualisation [SKM*09] can be seen showing an overview of all concepts from the two ontologies. An information landscape is a visualisation paradigm based on a geographic map metaphor, which conveys relatedness in the data through spatial proximity in the visualisation. Concepts are shown as dots with colours encoding ontology membership and spatial proximity between the dots encoding similarity between the concepts. By selecting a mapping in the table, the corresponding pair of concepts will be highlighted in the landscape. As related concepts from different ontologies are grouped together by the similarity layout algorithm, identification of regions rich with mapping candidates becomes easily possible: Identifying promising alignment candidates is as simple as finding dense regions containing dots in different colours, while regions dominated by one colour are

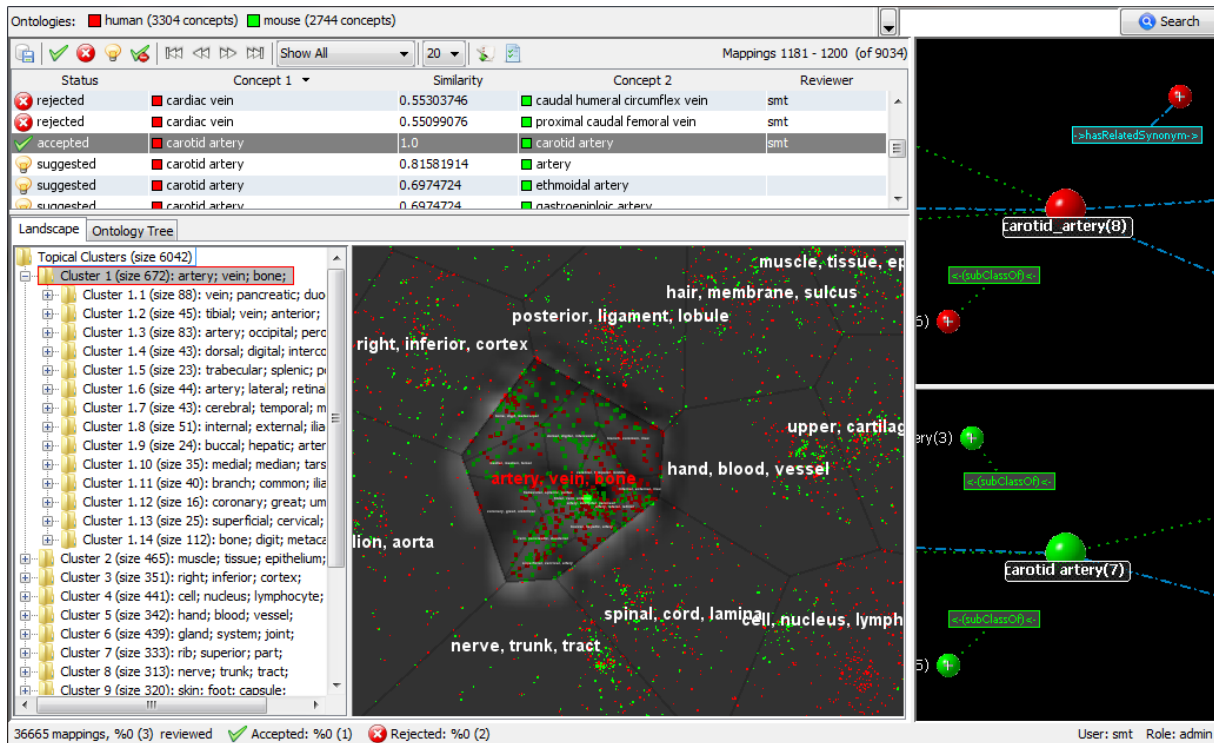


Figure 1: Semantic Mediation Tool user interface.

likely to contain no mappings (for example in the bottom-right corner of the landscape).

The landscape organises concepts from the two ontologies into a hierarchy of clusters which are represented as nested polygonal areas. Just like the concepts, cluster areas are also positioned in such a way that similar clusters are placed spatially close together. Each cluster is labelled by several keywords which were extracted as highest weight features from the underlying concept vectors. Labelled cluster areas are useful when aligning large ontologies, because the user can efficiently narrow down to the area of interest: Moving the mouse cursor over the keywords shows the area covered by the cluster (see "artery, vein, bone" cluster in Fig. 1). The user can navigate deeper in the cluster hierarchy by choosing a suitable cluster and clicking on its keywords, which reveals its sub-clusters. Alternatively, free navigation using zooming (mouse wheel) and panning (mouse drag) is also supported. Once the user has narrowed down to the area of interest, lasso- and/or cluster-selection can be used to select a particular group of concepts (highlighted in Fig. 1). On selection, the mapping table is updated to show only mappings containing the selected concepts.

In a classical information landscape hills will arise where the density of visualised items is large, visually emphasising

areas with agglomerations of similar items. To highlight the areas being reviewed by the user and separate them visually from areas which are currently not in the focus of interest, we have modified this concept so that only selected concepts, i.e. those for which mappings are currently shown in the mapping table, will contribute to the height of the hills.

The capability provided by the landscape to explore, narrow down and filter is crucial when the number of concepts and mappings is large, enabling the user to identify and focus only on relevant portions of the concept space. Note that, as users may be accustomed to navigating hierarchies using a tree widget, providing an additional tree view showing the cluster hierarchy (on the left from the landscape) proved to be beneficial to the users [GKS*04].

5.3. Ontology Browser

When deciding whether to accept or reject a mapping the user might need additional information on the corresponding pair of concepts. This information is provided visually by a pair of ontology browsers (on right in Fig. 1). When a pair of concepts is selected in the mapping table, each concept, together with the triples containing the concept, will be displayed in the corresponding ontology browser. Literals are shown with a gray 'information' icon, while predicates are

displayed as links with the name of the predicate labelling the link.

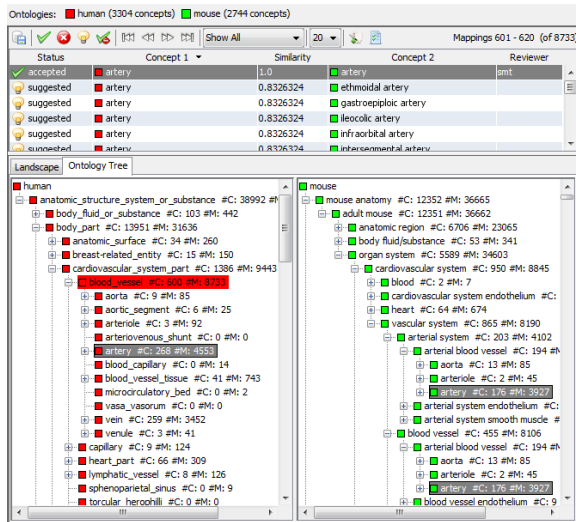


Figure 2: Trees showing ontology class hierarchies.

5.4. Ontology Trees

As an alternative to the landscape view, SMT provides a pair of tree views for navigation in the class hierarchies of both ontologies (see Fig. 2). Note that, while the hierarchy in the landscape view is calculated by the clustering algorithm, the ontology class hierarchies are typically man-made. The root element of each tree contains the name of the ontology. Besides the name of the concept, each concept label contains the number of sub-concepts in its sub-tree ($\#C$) and the number of mappings for itself and its sub-concepts ($\#M$). Narrowing down and filtering of the mappings is supported through navigation and selection. By selecting a concept (Fig. 2: "blood vessel", highlighted in red) the mapping table is updated to show only mappings containing the selected concept or its sub-concepts. Also, when a user selects a mapping in the mapping table, the trees will expand and the concepts will be highlighted (grey background). Note that due to multiple inheritance in the class hierarchy and flattening of the graph structure, it is likely that multiple branches in each tree will be expanded.

6. User Evaluation

We performed a user evaluation of the SMT to find out how our visual tool performs for tasks relevant to ontology alignment. In this evaluation we focused on comparing the usability of two visual representations of the concept space: the information landscape and the ontology trees. Since tree views are standard components in most operating systems, users will certainly be very familiar with them. However, due to

multiple inheritance in the class hierarchy and flattening of the ontology graph structure into a tree structure, concepts will be occurring multiple times in different sub-trees. We expect nodes occurring multiple times in different sub-trees to be counterintuitive for users, which might lead to slower interaction with the ontology trees, for example due to increased necessity for scrolling. Thus, our first hypothesis for the user evaluation, which focuses on navigational tasks involving exploration and navigation to a particular concept, is as follows:

H1 Users perform equally well with both representations in navigational tasks.

Information landscapes have been shown to be a useful representation for getting an overview of large data sets and narrowing down to the area of interest [SKM*09]. Thus, our second hypothesis for the user evaluation was the following:

H2 Users perform better with the information landscape in tasks involving narrowing-down and filtering.

During the evaluation we further wanted to find out users' preferences and collect general user feedback to the proposed tool.

6.1. Design

We used a within-subject design with the independent variable being the visual representation of the concept space with two different levels: information landscape and ontology trees. We measured task completion times and task completion success rates. Task completion time is measured as the time the user required from reading the task until the completion of the task or until the timeout for the task was reached. A task is counted as successfully solved if the desired result was achieved within the maximum time limit. The time limit was identified with three pilot-user tests and was set to one minute for the easy (first four) tasks and five minutes for the complex (last four) tasks.

6.2. Procedure

Figure 3 gives an overview of the evaluation procedure. Every participant was given an introduction to ontologies and to the goals of ontology alignment. Then, the participants were asked to fill out a demographic questionnaire. After that participants were introduced to the application and had the possibility to try it out and ask questions. Then each participant performed 8 tasks, where the last 4 tasks were performed twice, once with the information landscape and once with the ontology trees. The first four tasks were designed to be simple with the purpose to familiarise the participants with the terminology and the problem of ontology alignment. These initial four tasks, which were only performed using the mapping table, are as follows: **T1** finding the total number of mappings, **T2** identifying concepts for a first

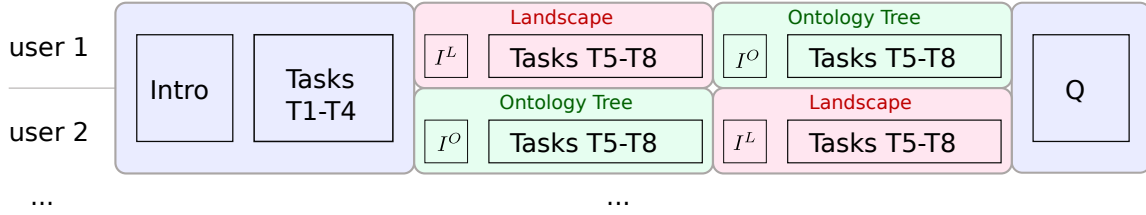


Figure 3: User evaluation procedure: After a general Introduction (I), all users performed tasks T1-T4. Tasks T5-T8 were performed twice by all users, once using the information landscape and once with the ontology trees. The sequence of the visual representations was altered for subsequent users (I^O - introduction to the Ontology Trees, I^L - introduction to information landscape). Finally, users filled out a questionnaire (Q).

mapping, **T3**) confirming the highest-, and **T4**) rejecting the lowest-confidence mapping for a given concept.

Each of the next four tasks were performed once with the information landscape and once with the ontology trees. The sequence of the test conditions was altered with each participant. Before beginning with the tasks the respective visual representation (information landscape or ontology trees) and available interaction mechanisms were explained in detail. The user had the possibility to try the interaction mechanisms and ask questions. Each user started testing with exactly the same state of the system for each task. Tasks T5-T8 were defined as follows:

- T5** Guided navigation to a concept and mapping counting,
- T6** Non-guided navigation and mapping review,
- T7** Overview, narrowing-down and elimination of non-relevant mapping subsets,
- T8** Narrowing-down and selection of a relevant mapping subset.

Tasks **T5** and **T6** were designed to test hypothesis **H1** (navigability), while tasks **T7** and **T8** were designed to test hypothesis **H2** (narrowing-down and filtering). After performing the tasks the participants filled out a questionnaire. The questionnaire contained questions about how the participants perceived the interaction with the mapping table, the information landscape and the ontology trees in terms of intuitiveness and ease of use. We used a 5-point Likert scale for these questions. Further, the participants stated which visual representation they preferred for tasks **T5** to **T8**. The questionnaire also contained a comments section about what the participants particularly liked and disliked, and what they think should be improved.

6.3. Test Material

For evaluating the SMT we used two standard ontologies from the *Ontology Alignment Evaluation Initiative* [OAE11], namely the **human** ontology consisting of 2744 concepts and **mouse** ontology containing 3304 concepts. For evaluation we used only the unsupervised learning-based alignment algorithm, the reason being that

alignment results and the similarity layout used by the landscape would be more in accord, since they were generated based on the same cluster hierarchy. The alignment algorithm generated 36.665 mapping suggestions, with multiple suggestions per concept enabled and the similarity threshold set to a low value in order to achieve a high recall.

As already mentioned, the cluster hierarchy shown in the information landscape is generated automatically based only on concept similarities. In contrast to that, the tree view displays the inherent hierarchical structure of an ontology by using the existing *subClassOf* relation and *UNDEFINED_part_of* restriction.

6.4. Participants and Environment

For the user evaluation we managed to recruit 15 volunteers, 13 males and 2 females. The age of the participants ranged from 24 to 40 years, with an average of 30.5 years. All participants were experienced computer users. One third (5) of the participants had little or no experience with ontologies and two thirds (10) were familiar or very familiar with ontologies. The participants were tested in a calm environment without noise distractions. The task was performed on a Dell Latitude e650 notebook running Windows XP Professional. The notebook was equipped with an Intel Core Duo 2.26 GHz, 3 GB RAM, a USB mouse and an external keyboard. An external 22 inch display with a resolution of 1680 x 1050 pixels was used.

7. Results of User Evaluation

In this section we present and discuss the results of time measurements, the questionnaire answers, and collected user suggestions.

7.1. Measured Performance

Table 1 shows the mean and standard deviation of the task completion times for tasks T5 to T8. We omit results for tasks T1 to T4, because the goal for those tasks was primarily to make participants familiar with the application. Only 12

out of 15 participants were able to complete task T7 within the given time limit, the results in Table 1 were calculated by omitting these users from the calculation for this task. We tested on equal means with Wilcoxon rank sum test for unpaired samples. The null hypothesis for the test was that the means are equal, we set $\alpha = .05$. We found a statistical difference ($\alpha = 0.05$) between the two conditions for task T5, T6, and T8. We found a statistical difference for task T7, which means users performed slower using the landscape in this task. However, all users were able to solve this task using the landscape within the given time limit. In contrast, 20% of the users were not able to solve this task at all using the ontology trees.

Task	Completion time [sec]		Success rate [%]	
	L	OT	L	OT
T5	61 ± 37	47 ± 15	100	100
T6	95 ± 37	89 ± 33	100	100
T7	179 ± 73	98 ± 55 ¹	100	80
T8	105 ± 57	112 ± 50	100	100

¹ the three participants who did not complete the task were omitted

Table 1: Comparing task completion times for tasks T5 to T8 for landscape (L) and ontology trees (OT). Showing mean and standard deviation. Statistically significant differences are marked bold.

We further were interested, if participants being slower using one visual representation are also slower using the other one. Thus, for Task T5 - T8 we calculated Pearson’s correlation coefficient ρ between the completion time participants achieved with the ontology trees and the completion time participants achieved using the landscape. The correlations ranged from 0.11 (T7) to 0.53 (T6). Given the limited amount of samples we conclude that there is no correlation between completion times using the different interfaces.

Summing up, on our user sample we found no difference in user performance between the information landscape and the ontology trees for navigational tasks (H1). Users performed not better with the information landscape in tasks involving narrowing-down and filtering (H2), users were able to solve all tasks with the information landscape, but not with the ontology trees.

7.2. Quantitative Evaluation of the Questionnaire

As mentioned before, each participant filled out a questionnaire after executing the test tasks. We report which visual representation users preferred for which task and how they perceived the interaction with the representation.

Table 2 summarises how participants perceived the visual representations, and Figure 4 shows which one was preferred by the participants for which task. As can be seen in the table, participants found the ontology trees more intuitive and

Property	Landscape	Ontology Trees
Intuitiveness	4.1 ± 0.8	3.8 ± 1.1
Ease-of-use	4.2 ± 0.6	3.3 ± 0.9

Table 2: User ratings for intuitiveness and ease-of-use of the landscape and the ontology trees on a five point scale where one is the best and five the worst value. Showing mean and standard deviation. Better (smaller) values are marked bold.

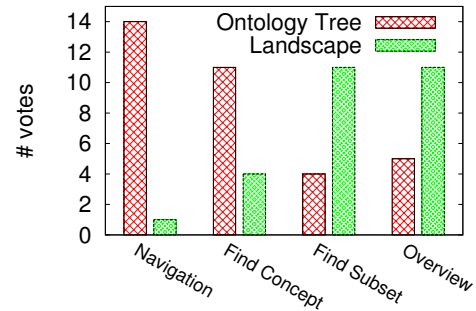


Figure 4: User voting: Preferred visual representation for tasks relevant to visual ontology alignment.

easier to use than the information landscape. However, Figure 4 shows a split picture. On the one hand, participants preferred the ontology trees for navigation and finding single concepts. On the other hand, for selecting relevant subsets of ontologies and getting an overview over the concept space, the landscape was the participants’ representation of choice.

Summing up, we found out, that participants found the ontology trees more intuitive, easier to use and better suited for navigation tasks, but clearly preferred the landscape for overview and subset selection.

7.3. Qualitative Evaluation of the Questionnaire

Here we give a report on comments provided by test users including positive and negative feedback as well as suggestions for improvement. Due to the large number of comments, we provide insights into the overall distribution of comments and give a summary of the most common issues. A detailed discussion would be outside the scope of this paper and will be provided in a future report. All 15 participants provided at least one comment. The comments could be grouped into five categories which are listed in Table 3. Some of the comments addressed multiple categories and/or multiple visual components. For the following analysis we counted such comments once for each addressed category or component, resulting in 86 comments in total. We evaluated the type of the comments, identifying each comment either as being positive, being negative or being a suggestion. Figure 5 gives an overview of the number of comments, their categories and their types for each component.

Category	Description
Features	Requests for new features, suggestions for improving existing ones
Filtering	Comments on filtering mappings in the mapping table
Navigation	Comments on navigation in ontology trees and landscape
Visual Appearance	Comments about look and feel of different components of the application
Technical Issues	Comments about the technical issues, such as response times or click accuracy

Table 3: Categories of participants' comments.

Majority of the comments (27) were about navigation, most of them (12) being about navigation in the ontology trees. Distribution of positive and negative comments on navigation in ontology trees was equal, negative comments arising mostly due to multiple branches being expanded on concept selection in the mapping table. Navigation comments for the landscape (7) are slightly on the negative side (3 positive vs. 4 negative), with negative comments addressing loss of context when zooming and technical issues concerning the mouse interaction. Comments on navigation in the mapping table was predominantly negative (2 positive vs. 4 negative comments), mostly due to complexity arising from the combination of sorting, paging and selection.

In terms of components, the landscape occurred in most comments (32), about 25% of the comments covering its visual appearance. The visual appearance of the information landscape received strongly positive comments (7 positive vs. 3 negative comments). In total, the landscape (ratio positive to negative comments: 13/11) was perceived more positively than ontology trees (ratio positive to negative comments: 3/7).

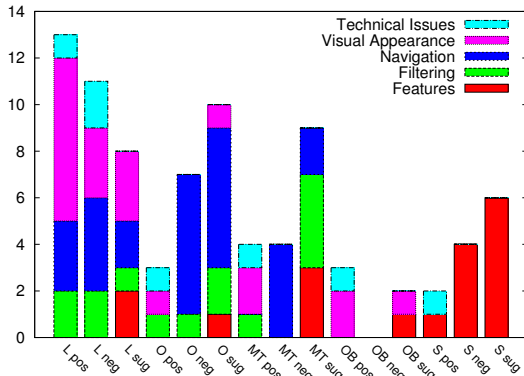


Figure 5: Positive (pos), negative comments (neg) and suggestions (sug) by component and category. OT - Ontology Trees, L - Landscape, MT - Mapping Table, OB - Ontology Browser, S - Search.

Summing up, the user rated the landscape more positively than ontology trees, but requested improvements in navigation and fixes for mouse interactions.

8. Conclusion

We presented a semi-automatic visual system for ontology alignment, in which algorithmic methods are used to compute an initial set of mapping suggestions, which are then reviewed by experts using a visual user interface. In a user evaluation we compared the two visual key components, the information landscape and ontology trees, for tasks relevant to visual ontology alignment.

We found no difference in task completion times between the landscape and the ontology trees for navigational tasks. However, landscape appears to be slightly slower, but the results are not statistically significant on the tested user sample. However, the information landscape is better suited for narrowing down, filtering and selection in terms of task completion success rate. On the basis of the comments to the questionnaire users preferred the landscape for overview tasks. However, users would prefer the ontology trees for navigational tasks, but as indicated by user comments this might change if users were provided a better training for the unfamiliar landscape interface.

Concerning our future work, in the short term we will be fixing a number of minor but annoying technical and usability issues, especially those affecting the landscape visualisation. We will also address the most pressing feature requests for improving mapping selection and navigation. In the following, we plan to focus on evaluation and tuning of the alignment algorithms. As a challenging goal for the future, we envision a system where user feedback provided in visual form will be utilised to adapt the model, improve the algorithm performance, and show the adapted results within a short (possibly near real-time) time interval. Finally, after having compared our visual tool against a common tree-based interface, we will compare SMT performance to other visual ontology alignment systems, such as those mentioned in Section 2.

Acknowledgements The Know-Center is funded within the Austrian COMET Program – Competence Centers for Excellent Technologies – under the auspices of the Austrian Ministry of Transport, Innovation and Technology, the Austrian Ministry of Economics and Labor and by the State of Styria. COMET is managed by the Austrian Research Promotion Agency (FFG). MIMOS Berhad is funded by the Malaysian government through the Ministry of Science, Technology and Innovation (MOSTI).

References

- [ADMR05] AUMUELLER D., DO H.-H., MASSMANN S., RAHM E.: Schema and ontology matching with com++ , June 2005. In *Proceedings of the ACM SIGMOD*. 10
- [Aur91] AURENHAMMER F.: Voronoi diagrams - a survey of a fundamental geometric data structure. *ACM Computing Surveys* (1991), 345–405. 11
- [CDD*02] CARROLL J. J., DICKINSON I., DOLLIN C., REYNOLDS D., SEABORNE A., WILKINSON K.: Jena - A Semantic Web Framework for Java. <http://jena.sourceforge.net/index.html>, 2002. 10
- [CSMB07] CRUZ I., SUNNA W., MAKAR N., BATHALA S.: A visual tool for ontology alignment to enable geospatial interoperability. *Journal of Visual Languages and Computing* (2007), 230–254. 10
- [dSDdMR06] DE SOUZA K. X. S., DAVIS J., DE MEDEIROS E., ROBERTO S.: Aligning ontologies, evaluating concept similarities and visualizing results. *Journal on Data Semantics V* (2006), 211–236. 10
- [EBJ06] EBBELS T. M. D., BUXTON B. F., JONES D. T.: springscape: visualisation of microarray and contextual bioinformatic data using spring embedding and an 'information landscape'. *Bioinformatics* 22 (2006), 99–107. 10
- [ELBB*04] EUZENAT J., LE BACH T., BARRASA J., BOUQUET P., DE BO J., DIENG R., EHRIG M., HAUSWIRTH M., JARRAR M., LARA R., MAYNARD D., NAPOLI A., STAMOU G., STUCKENSCHMIDT H., SHVAIKO P., TESSARIS S., VAN ACKER S., ZAIHRAYEU I.: State of the art on ontology alignment, 2004. Deliverable of the Knowledge Web Project (IST-2004-507482), Knowledge Web Consortium. 9
- [ES07] EUZENAT J., SHVAIKO P.: Ontology matching, 2007. 9
- [FBG09] FALCONER S., BULL R., GRAMMEL L. AND STOREY M.-A.: Creating visualizations through ontology mapping, March 2009. In *Proceedings of the 2nd International Workshop on Ontology Alignment and Visualization*. 10
- [FNS06] FALCONER S., NOY N., STOREY M.-A.: Towards understanding the needs of cognitive support for ontology mapping. In *Proceedings of the Ontology Matching Workshop (5th International Semantic Web Conference)* (2006), pp. 25–36. 10
- [FNS07] FALCONER S., NOY N., STOREY M.-A.: Ontology mapping - a user survey. In *Proceedings of the Workshop on Ontology Matching (OM2007) at ISWC/ASWC2007* (2007), pp. 113–125. 10
- [FR91] FRUCHTERMAN T., REINGOLD E.: Graph drawing by force-directed placement. *Software – Practice & Experience* (Wiley) (1991), 1129–1164. 11
- [FS07] FALCONER S., STOREY M.-A.: A cognitive support framework for ontology mapping, November 2007. In *Proceedings of International Semantic Web Conference*. 10
- [GKS*04] GRANITZER M., KIENREICH W., SABOL V., ANDREWS K., KLIEBER W.: Evaluating a system for interactive exploration of large, hierarchically structured document repositories. In *Proceedings of the IEEE Symposium on Information Visualization (InfoVis '04)* (2004), IEEE Computer Society, pp. 127–134. 13
- [GS08] GAL A., SHVAIKO P.: Advances in ontology matching. In *Advances in Web Semantics I: Ontologies, Web Services and Applied Semantic Web*. Springer, 2008, pp. 176–198. 9
- [GSK*10] GRANITZER M., SABOL V., KOW W. O., LUKOSE D., TOCHTERMANN K.: Ontology alignment - a survey with focus on visually supported semi-automatic techniques. *Future Internet* 2, 3 (2010), 238–258. 10, 12
- [KD08] KOLLI R., DOSHI P.: Optima: Tool for ontology alignment with application to semantic reconciliation of sensor metadata for publication in sensormap, August 2008. In *Proceedings of the second IEEE International Conference on Semantic Computing*. 10
- [KL08] KOTIS M., LANZENBERGER M.: Ontology matching: Status and challenges. *IEEE Intelligent Systems* 23 (2008), 84–85. 9
- [KSG*11] KOW W. O., SABOL V., GRANITZER M., KIENREICH W., LUKOSE D.: A visual soa-based ontology alignment tool. In *Proceedings of the Sixth International Workshop on Ontology Matching (OM 2011)* (2011). 10
- [KSM*09] KLIEBER W., SABOL V., MUHR M., KERN R., ÖTTL G., GRANITZER M.: Knowledge discovery using the knowminer framework, iads. In *IADIS International Conference Information Systems* (2009), pp. 307–314. 11
- [LS06] LANZENBERGER M., SAMPSON J.: Alviz - a tool for visual ontology alignment. In *In IV '06: Proceedings of the conference on Information Visualization* (2006), IEEE Computer Society, pp. 430–440. 10
- [LSR*08] LANZENBERGER M., SAMPSON J., RESTER M., NAUDET Y., LATOUR T.: Visual ontology alignment for knowledge sharing and reuse. *J. Knowledge Management* 12, 6 (2008), 102–120. 10
- [MSG10] MUHR M., SABOL V., GRANITZER M.: Scalable recursive top-down hierarchical clustering approach with implicit model selection for textual data sets. In *7th International Workshop on Text-based Information Retrieval in Proceedings of 21th International Conference on Database and Expert Systems Applications (DEXA 10)* (2010). 11
- [NM03] NOY N., MUSEN M.: The prompt suite: Interactive tools for ontology merging and mapping. *International Journal of Human Computer Studies* 59 (2003), 983–1024. 10
- [OAE11] Ontology Alignment Evaluation Initiative. [//oaei.ontologymatching.org](http://oaei.ontologymatching.org), 2011. 9, 12, 15
- [PM00] PELLEG D., MOORE A.: X-means: Extending k-means with efficient estimation of the number of clusters, 2000. 11
- [Rah11] RAHM E.: Towards large-scale schema and ontology matching. In *Schema Matching and Mapping*. Springer, 2011, pp. 3–27. 9
- [Shn91] SHNEIDERMAN B.: Tree visualization with tree-maps: A 2-d space-filling approach. *ACM Transactions on Graphics* 11 (1991), 92–99. 10
- [SKM*09] SABOL V., KIENREICH W., MUHR M., KLIEBER W., GRANITZER M.: Visual knowledge discovery in dynamic enterprise text repositories, 2009. 12, 14
- [TC05] THOMAS J., COOK K. (Eds.): *Illuminating the Path: The Research and Development Agenda for Visual Analytics*. IEEE Computer Society, August 2005. National Visualization and Analytics Center. 10
- [UGM07] UDREA O., GETOOR L., MILLER J.: Homer: Ontology alignment visualization and analysis. In *6th International and 2nd Asian Semantic Web Conference (ISWC2007+ASWC2007)* (November 2007), pp. 111–112. 10
- [Uni10] UNIVERSITY P.: About wordnet. <http://wordnet.princeton.edu>, 2010. 11
- [W3C09] W3C: Allegrograph rdfstore web 3.0's database. <http://www.franz.com/agraph/allegrograph/>, September 2009. 10
- [WP94] WU Z., PALMER M.: Verb semantics and lexical selection, 1994. *Proceedings 32nd Annual Meeting of the Association for Computational Linguistics (ACL)*. 11