NIH-PA Author Manuscript

NIH-PA Author Manuscript

NIH-PA Author Manuscript

NIH Public Access
Author Manuscript
Int J Semant Web Inf Syst. Author manuscript; available in PMC 2014 January 24.

Published in final edited form as:
Int J Semant Web Inf Syst. 2013 ; 9(1): 45–78. doi:10.4018/jswis.2013010103.

PragmatiX: An Interactive Tool for Visualizing the Creation Process Behind Collaboratively Engineered Ontologies

Simon Walk,
Knowledge Management Institute, Graz University of Technology, Inffeldgasse 21a/II, 8010 Graz

Jan Pöschko,
Knowledge Management Institute, Graz University of Technology, Inffeldgasse 21a/II, 8010 Graz

Markus Strohmaier,
Knowledge Management Institute, Graz University of Technology, Inffeldgasse 21a/II, 8010 Graz

Keith Andrews,
Institute for Information Systems and Computer Media, Graz University of Technology,
Inffeldgasse 16c, 8010 Graz

Tania Tudorache,
Stanford Center for Biomedical Informatics Research, Stanford University, 1265 Welch Road,
Stanford, CA 94305-5479, USA

Natalya F. Noy,
Stanford Center for Biomedical Informatics Research, Stanford University, 1265 Welch Road,
Stanford, CA 94305-5479, USA

Csongor Nyulas, and
Stanford Center for Biomedical Informatics Research, Stanford University, 1265 Welch Road,
Stanford, CA 94305-5479, USA

Mark A. Musen
Stanford Center for Biomedical Informatics Research, Stanford University, 1265 Welch Road,
Stanford, CA 94305-5479, USA

Abstract

With the emergence of tools for collaborative ontology engineering, more and more data about the creation process behind collaborative construction of ontologies is becoming available. Today, collaborative ontology engineering tools such as Collaborative Protégé offer rich and structured logs of changes, thereby opening up new challenges and opportunities to study and analyze the creation of collaboratively constructed ontologies. While there exists a plethora of visualization tools for ontologies, they have primarily been built to visualize aspects of the final product (the ontology) and not the collaborative processes behind construction (e.g. the changes made by contributors over time). To the best of our knowledge, there exists no ontology visualization tool today that focuses primarily on visualizing the history behind collaboratively constructed ontologies. Since the ontology engineering processes can influence the quality of the final ontology, we believe that visualizing process data represents an important stepping-stone towards better understanding of managing the collaborative construction of ontologies in the future. In this *application paper*, we present a tool – *PragmatiX* – which taps into structured change logs

Note to reviewers:
We invite the reviewers to browse the full Heuristic Evaluation Report for this paper, including evaluation screenshots and other information, online at:
Heuristic Evaluation Report, http://heuristiceval.simonwalk.at/he.html

NIH-PA Author Manuscript NIH-PA Author Manuscript NIH-PA Author Manuscript

provided by tools such as Collaborative Protégé to visualize various pragmatic aspects of collaborative ontology engineering. The tool is aimed at managers and leaders of collaborative ontology engineering projects to help them in monitoring progress, in exploring issues and problems, and in tracking quality-related issues such as overrides and coordination among contributors. The paper makes the following contributions: (i) we present *PragmatiX*, a tool for visualizing the creation process behind collaboratively constructed ontologies (ii) we illustrate the functionality and generality of the tool by applying it to structured logs of changes of two large collaborative ontology-engineering projects and (iii) we conduct a heuristic evaluation of the tool with domain experts to uncover early design challenges and opportunities for improvement. Finally, we hope that this work sparks a new line of research on visualization tools for collaborative ontology engineering projects.

Keywords

Collaborative Ontology Engineering; pragmatic analysis; ontology monitoring; ontology engineering visualization; ontology evaluation; ontology tool

INTRODUCTION

While collaboration, negotiation, and consensus represent an integral part of ontology engineering processes, it is only recently that disciplined tools and infrastructure for *collaborative* ontology engineering have emerged. Tools such as Collaborative Protégé (Tudorache, Noy, Tu, & Musen, 2008) not only provide an infrastructure for collaboration and coordination, but also provide a structured log of all ontological changes, which users have made via the tool. These logs can, for example, include records of concepts added, properties changed, or relationships qualified. In aggregation, such logs can essentially capture the entire evolution of an ontology from its inception to its final stages on a very fine-grained level. At the same time, the availability of fine-grained logs poses new challenges and opportunities for studying and analyzing the history of collaborative ontology engineering projects. While there exists a plethora of visualization tools for ontologies, they have primarily been built to visualize aspects of the final product (the ontology) and not the collaborative processes behind construction (e.g. the changes made by contributors over time). To the best of our knowledge, there exists no ontology visualization tool today that focuses primarily on visualizing the creation processes behind collaboratively constructed ontologies.

This *application* paper sets out to present a visualization tool that primarily focuses on visualizing pragmatic aspects of collaborative ontology engineering, i.e. the social processes that yield collaboratively constructed ontologies. We present a tool – *PragmatiX* – that taps into structured log of changes provided by tools such as Collaborative Protégé and visualizes them via network-based and other kinds of visualizations. The tool is aimed at managers and leaders of collaborative ontology engineering projects to help them in monitoring progress, exploring issues and problems, and tracking quality-related issues such as overrides and coordination among contributors. *PragmatiX* is the successor of iCAT Analytics (Pöschko, Strohmaier, Tudorache, Noy, & Musen, 2012) and provides additional functionality such as the heat-map (as described in Section *Concept Network Visualization)*, the possibility of importing multiple data sets into one instance of our tool, the support for multi-language data sets (see Section *Category and Author Views*) as well as various statistical overview pages such as the dashboards (see Section *Dashboard*). Additionally, a heuristic evaluation has been performed on our tool, providing interesting results for future work.

NIH-PA Author Manuscript NIH-PA Author Manuscript NIH-PA Author Manuscript

Our initial prototype demonstrates its capabilities by tapping into change-logs produced by variants of Collaborative Protégé, where changes and notes as well as comments on changes are represented in the Change and Annotation Ontology (ChAO) (Noy, Chugh, Liu, & Musen, 2006). Because several large collaborative ontology-engineering projects in the bio-medical domain use Collaborative Protégé (and its derivatives) for tool support, we have access to change-log data from a series of different projects. For example, the International Classification of Diseases (ICD-11) project uses WebProtégé, a Web version of Protégé that is built on the collaborative framework of Collaborative Protégé, to collaboratively engineer a bio-medical ontology consisting of more than 30,000 concepts (Tudorache, Falconer, Nyulas, Noy, & Musen, 2010). Almost all changes to this ontology have been captured and are available for further analysis. The International Classification of Traditional Medicine (ICTM) ontology represents another example, where a sufficiently large record of changes is available. In this paper, we will use data from these two projects to demonstrate the general applicability of our tool for visualizing pragmatic aspects of collaborative (ontology-) engineering projects. While the illustrations in this paper are limited to these two projects, there is nothing in our implementation, which prevents other collaborative ontology engineering projects (e.g. outside the bio-medical domain) being visualized with our tool in a similar manner, given that data about the creation process is available in a structured form (see section *Implementation*).

Our application paper makes the following contributions: (i) we present the new and extended version of iCAT Analytics, called *PragmatiX*, which allows for visualizing the creation process behind collaboratively constructed ontologies (ii) we illustrate the utility and generality of the tool by applying it to structured change-logs of two large collaborative ontology-engineering projects and (iii) we conduct a heuristic evaluation of the tool with domain experts to uncover early design challenges and opportunities for improvement. Our research is relevant for managers of collaborative ontology engineering efforts aiming to analyze and visualize the social dynamics of the development process.

This paper is structured as follows: The *Related Work* section provides a detailed overview of existing and relevant published work, which has influenced this paper. In the *Materials and Methods* section we describe the resources used by *PragmatiX*, such as various data sets and their structured logs of changes as well as the algorithms used to calculate the features displayed in the different views. In section *The PragmatiX Visualization Tool*, we describe *PragmatiX* itself in greater detail, including all the visualizations and views provided by the tool. The results of a formative evaluation of *PragmatiX* are presented in the *Evaluation* section; the benefits and limitations of *PragmatiX* are outlined in the *Discussion* section. The paper closes with concluding remarks in section *Conclusions*, where we additionally address possible future work.

RELATED WORK

The following areas of research are relevant to our work: Collaborative Ontology Engineering, Collaborative Ontology Engineering Tools, and Ontology Visualization Tools.

Collaborative Ontology-Engineering

The field of ontology engineering covers many different topics ranging from best practices for creating ontologies (Cristani & Cuel, 2005; Noy & McGuinness, 2001; Spyns, Meersman & Jarrar, 2002), identifying and implementing semi-automatic processes to create ontologies from different resources such as plain text (Maedche & Staab, 2000) to the task of ontology-evaluation (Brank, Grobelnik & Mladenic, 2005) in order to determine and quantify the quality of an ontology, for example with respect to its intended use-case.

NIH-PA Author Manuscript

NIH-PA Author Manuscript

NIH-PA Author Manuscript

In contrast to traditional ontology engineering, the task of collaboratively developing and engineering an ontology represents an emergent field of research with new problems, risks and challenges.

For example, Noy & Tudorache (2008) and Falconer, Tudorache, & Noy (2011) focus on identifying, defining and surveying requirements for collaborative ontology-engineering applications. Their work demonstrates that an analysis of change-logs of collaboratively engineered ontologies allows users to be grouped according to their change behavior. Pöschko et al. (2012) have shown that analyzing the structured log of changes in collaborative ontology engineering projects using *iCAT Analytics*, the predecessor to *PragmatiX*, yields interesting results, such as how work is distributed among authors or which areas of the ontology already received a large amount of contributions, which can be used to enhance the collaborative engineering process and to help encourage users to contribute. In contrast to iCAT Analytics, PragmatiX allows the import and visualization of multiple data sets in one instance. It also provides additional functionalities such as various statistical overview pages (e.g. the heat-map as described in Section *Concept Network Visualization*, or the dashboards in Section *Dashboard*). Additionally, a heuristic evaluation has been performed on *PragmatiX*, providing interesting results for future work.

Collaborative Ontology-Engineering Tools

Many collaboratively engineered ontologies, such as the Gene Ontology (GO) (Harris et al., 2004), the National Cancer Institute Thesaurus (NCIt) (Golbeck et al., 2003), the International Classification of Diseases revision 11 (ICD-11) and the International Classification of Traditional Medicine (ICTM) (Tudorache et al., 2010), are created using tools that provide special methods and functionality to help users collaborate. This special functionality often includes mechanisms to comment single concepts, to engage in discussion, and to justify changes and design decisions, all of which support collaboration among ontology editors.

A large variety of ontology-engineering tools, such as OntoEdit (Sure et al., 2002), semantically extended Wikis, such as Wiki@nt (Bao & Honavar, 2004) and OntoWiki (Auer, Dietzold, & Riechert, 2006), or Collaborative Protégé and WebProtégé (Noy & Tudorache, 2008; Tudorache, Noy, et. al., 2008) provide special functionality supporting users in reaching consent and avoiding conflicting changes by actively encouraging collaboration.

Both of the collaborative ontology engineering projects that we use in this paper were developed using the web-based tools iCAT (see Figure 1) and iCAT TM, two very similar, customized versions of WebProtégé. The most important feature of WebProtégé and its derivatives for this paper is the fact that it provides a very detailed and fine-grained structured log of changes of the ontology, which can be used to analyze the creation processes in addition to the collaboratively constructed ontologies.

iCAT and iCAT TM both offer extensive collaborative features, providing authors not only with the ability to conduct collaborative work but also to engage in threaded discussions, to facilitate collaborative decision making.

Ontology Visualization Tools

The domain of ontology visualization covers a large set of applications providing various graphical representations for ontologies, which range from simple indented lists or trees, 2-dimensional graph representations to very sophisticated 3-dimensional layouts.

NIH-PA Author Manuscript

NIH-PA Author Manuscript

NIH-PA Author Manuscript

For example, Jambalaya (Storey et al., 2002b) was developed as a plug-in for an earlier version of Protégé. It uses a visualization technique called SHriMP (Simple Hierarchical Multi-Perspective) (Storey et al., 2002a), which supports the concept of interchangeable nested views representing an ontology in 2-dimensional space.

OntoViz (Singh et al., 2006), another Protégé visualization plug-in, on the other hand represents an ontology as a 2-dimensional graph, using the Graphviz (Ellson, Gansner, Koutsofios, North, & Woodhull, 2001) library. In OntoViz, every node represents a class or an instance, which in turn can display its name and some or all of its (inherited) properties and roles. Every edge represents a relationship between classes, instances, or both.

OWLViz (Horridge, 2012), TGViz (Alani, 2003), and OntoGraf (Falconer, 2010) are ontology visualization plug-ins for Protégé, which are similar to OntoViz. They represent an ontology as a 2-dimensional graph, where each node represents either a class or an instance and every edge represents a relationship between two entities. However, instead of visualizing all property and role information, OWLViz and TGViz reduce visual clutter by providing detailed information for each entity only when it is selected. Additionally, they allow filtering for specific parts of the ontology.

AlViz (Lanzenberger & Sampson, 2006) is a tool that was specifically designed to visualize and augment the task of ontology-alignment, i.e. mapping the classes and instances of one ontology to the classes and instances of another ontology. Classes and instances are represented as nodes, which are colored according to the result of the alignment process.

In contrast, OntoRama (Eklund, Roberts, & Green, 2002) uses a hyperbolic-type layout to visualize an ontology. This approach emphasizes classes or instances in the center of the visualization, as they are assigned more space and display a higher level of detail, while nodes near the edges of the visualization are minimized and only display a low level of detail. OntoSphere 3D (Bosca, Bonino, & Pellegrino, 2005) uses different types of 3-dimensional visualization to support users in browsing and exploring the structure and complexity of an ontology.

However, all of these ontology visualization tools have their focus on visualizing an already created ontology or (parts of) a static snapshot of an ontology. The idea and overall objective for *PragmatiX* is to create a tool, which not only visualizes an ontology, but also visualizes and analyzes the creation process.

MATERIALS AND METHODS

In this section we will first describe all identified requirements including the target audience of *PragmatiX* as well as the data sets and the structured logs of changes, which were used in our analysis and evaluation.

User Research

PragmatiX was specifically designed to augment; support and help to enrich and enhance the work performed by different user-types and classes (Schreiber et al., 2000) of knowledge-based system development processes. We have grouped these roles into the following three groups: The **Administrative Personnel**, which is composed of *Knowledge Managers* and *Project Managers*. The **Engineering Staff**, which in turn is composed of *Knowledge Engineers* and *Analysts*, *Knowledge System Developers* and *Knowledge Providers*. The final group of users is called **Ontology Viewers**. They consist of *Knowledge Users* and *System Visitors*. These user-types differ with regard to their informational needs and overall goals and objectives, for example Knowledge Users and System Visitors are

NIH-PA Author Manuscript NIH-PA Author Manuscript NIH-PA Author Manuscript

mainly interested in gathering a quick characterization of the data set or the tool. The Engineering Staff is concerned about the correctness of the underlying data, while Administrative Personnel wants to track the progress of the project.

For the purpose of identifying informational needs as well as the goals and objectives of different user types, meetings with members of the Protégé-Team at Stanford Center for Biomedical Informatics Research and with Domain Experts from the World Health Organization (WHO) were held. In those meetings, requirements for the tool were elicited and discussed with the stakeholders (consisting of the team that develops Protégé and the team that is in charge of ICD-11 development) in an iterative manner.

Implementation

The majority of *PragmatiX* was written in Python using the web-framework Django. We use NetworkX (Hagberg, Schult, & Swart, 2008) for all network calculations and use Graphviz (Ellson et al., 2001) to pre-calculate the different network layouts (visualizations). The data sets were exported from iCAT and iCAT TM using their Java API and stored in a MySQL database.

To visualize the different network views we make use of a combination of JavaScript, AJAX calls, and JSON. The asynchronous JavaScript and XML calls are necessary to update the graph after user interactions. Most of the visualizations and analyses available in *PragmatiX* are pre-calculated, resulting in reasonable response times and relatively low server load. This is especially useful for the network visualizations, where all positions are pre-calculated using Graphviz, and stored in the database. This approach minimizes calculation and loading times, since all required information can be extracted directly from the database without having to invoke additional computational tasks on either the client or the server. Additionally, most values displayed in a pie or line chart are pre-calculated.

PragmatiX additionally provides mechanisms to import data in a specific input format (basic txt-files), where each line corresponds to one concept followed by a very limited set of attributes extracted directly from the ontology, separated by tab-stops. This set of attributes consists of (i) a unique concept id, (ii) a concept title or username, (iii) a concept definition or change message, (iv) the assigned communities of interest (user groups; only if available) of the concept and (v) a concept's assigned display status color code, depending on the type of concept.

The ChAO change-log provides detailed change-information such as the user who performed a change, the concept it is performed on and a detailed change-description such as “Moved class: R75.2b Niemann-Pick disease. Old parent: E75.2 Other sphingolipidosis, New parent: Sphingolipidosis”, allowing to omit the storage of additional attributes (other than the ones mentioned), as they can be automatically generated for each point in time by processing the change-descriptions of the structured logs of changes. Not having to know which properties are available for which concept additionally increases the generality of *PragmatiX*. All additional attributes (such as the features displayed in Table 2) are calculated after the initial import.

The used input format for *PragmatiX* can easily be reproduced independent from the original source, given a structured log of changes is provided which can be mapped onto the ontology. Using this convention, we were able to import and visualize all articles from the official (and freely available) Wikipedia change-data dumps, which are marked with an ICD-10 code, into *PragmatiX*. The extracted articles have been mapped to concepts while contributors in Wikipedia represent users in *PragmatiX*. Relationships (edges) have been extracted from the original ICD-10 ontology.

NIH-PA Author Manuscript

NIH-PA Author Manuscript

NIH-PA Author Manuscript

Once the data is available as textual files in the required format, PragmatiX provides SQL scripts that import the content into the database. In a next step, the pre-calculations, a python script provided by PragmatiX, have to be initiated, which are then automatically performed. Once the pre-calculations are done, the new data set has to be added to the configuration-file and is afterwards ready to be browsed. We are currently working on refining the import process to automate all necessary steps and provide a step-by-step guide for all steps that cannot be automated.

Once PragmatiX reaches a stable version, we will consider releasing it as Open Source Software.

Application

PragmatiX focuses on visualizing the creation process behind collaborative ontology-engineering projects that provide (i) structural and contextual information about the ontology and (ii) a structured log of changes (and notes) that allows mapping every logged action to a specific user and the affected concept(s).

We have applied *PragmatiX* to five different collaboratively engineered ontologies from the bio-medical domain. Due to limitations in space, in this paper we will demonstrate the application of *PragmatiX* to only two of the five available bio-medical ontologies, which were both constructed using variations of WebProtégé. The two projects are:

- *ICD-11*: The structured log of changes comprises *152,955* changes and *31,197* notes over an observation period of 24 months. The ontology itself consists of *33,714* concepts and *76* users that performed all the changes.
- *ICTM*: This data set is of a smaller dimension and only consists of *1,311* concepts with a total of *21* users that actively work on the ontology. The change log consists of *39,495* changes and *1,449* notes over an observation period of 10 months.

Even though both data sets are maintained by WHO and have been created using either Protégé or one of its derivatives it is important to note that *PragmatiX* can be adapted to support every collaborative ontology engineering project which exhibits a structured log of changes.

Evaluation

Formative usability evaluation is usually performed during interface development, in order to identify potential problems to be fixed in future releases. Two classic methods of formative evaluation are widely used in software development: Heuristic Evaluation (HE) (Nielsen & Mack, 1994) and Thinking Aloud (TA) (Barnum, 2010) testing. The former involves a small group of specialist evaluators who inspect an interface and use a list of heuristics, combined with their knowledge and experience to identify and classify potential problems.

The latter involves a small number of representative test users from the target user population, who talk out loud whilst performing representative tasks, thus providing insight into their thought process when problems occur. Summative evaluations (Rubin, Chisnell, & Spool, 2008) involve the objective measurement of performance metrics and statistical analysis and are often used to compare alternative designs or competing products.

Because of the early stage of development, we have limited our evaluation efforts, and concentrated on conducting a Heuristic Evaluation on *PragmatiX* with three ontology-engineering experts who investigated and explored our tool in sessions of 60-120 minutes.

NIH-PA Author Manuscript

NIH-PA Author Manuscript

NIH-PA Author Manuscript

We have gathered feedback about the utility and problems assigned to *PragmatiX*, which we will further discuss in our section *Evaluation*.

THE PragmatiX VISUALIZATION TOOL

PragmatiX represents an evolution of the iCAT Analytics tool (Pöschko et al., 2012) and goes beyond iCAT Analytics by (i) being applicable to collaboratively engineered ontologies in general (and not specific to a particular ontology) and (ii) by adding several new views and visualizations to its repertoire. PragmatiX provides several different ways to interact with the analyzed data sets, which will be described in this section in greater detail. Users can perform exploratory analyses using different kinds of visualizations including **three network visualizations, ranked overviews** and **detailed statistics views** for all concepts and users. To further accommodate the needs of *Administrative Personnel,* we extended the tool by implementing a *dashboard*, which lists rather general statistics that can be used to interpret and monitor the progress of the underlying ontology engineering process.

Tool Overview

PragmatiX provides several different views that allow for different types of interaction with the imported data sets (see Figure 2). These different views and network visualizations are listed in Figure 2 and consist of:

- The **concept network visualization** hierarchically visualizes the concepts of an ontology via *is-a relations* and simultaneously allows to further visually inspect/ explore conceptual features, such as the number of changes performed on every concept of the ontology.
- The **author network visualization** visualizes the relationships across users by identifying and quantifying commonly edited concepts (or collaboration).
- The **property network visualization** displays properties of concepts, and their *pragmatic* relationships with each other (e.g. what property was edited after what other property?)
- The **dashboard & community views** are used to visualize and list general statistics which support *Administrative Personnel* in monitoring the progress of the engineering process.
- The **statistics overviews** feature rankings of all concepts, authors and properties according to several different pre-calculated features.
- The **detailed statistics views** provide detailed information about the change-history of a single concept or a single author.

Concept Network Visualization

This network is used to visualize change data specific to hierarchical structures, relations and to the complexity of and between the concepts of an ontology. Every node represents a single concept in the ontology. Every edge represents *is-a* relationships between concepts. The color assigned to each concept represents the *display status* of the concept – a property of the ICD-11 ontology, which can be used to represent the current development state or progress of a concept. We have adopted the color codes from *iCAT* and visualize them if they are supported by the underlying imported collaboratively engineered data set. In the case of ICD-11 the colors have the following meanings:

- Gray: no display status assigned

NIH-PA Author Manuscript NIH-PA Author Manuscript NIH-PA Author Manuscript

- Red: the concept requires extensive work
- Yellow: the concept is worked on, but is not finished
- Blue: the concept is ready for subsequent phases

These color codes are usually assigned by managers or leaders of collaborative ontology engineering projects, but could also be assigned by other roles (e.g. editors) through other mechanisms (e.g. voting). The color codes provide a quick overview of the current state of the ontology, which is especially important for collaboratively engineered ontologies as they can help to minimize the difficulties of identifying concepts or areas of an ontology that still need work, without having to allocate additional resources for that task.

Additionally the concept network visualization allows users to decide what determines the diameter of each node by selecting from a set of conceptual features (see Table 1) that are used as weights. This allows *PragmatiX* to help answer a series of questions about the creation processes behind a particular collaboratively engineered ontology.

In addition, *PragmatiX* allows limiting the visualization of the concept network to only display the set of concepts a specific user has edited, weighted according to the features listed in Table 1. The network itself is created analogously to all other network visualizations, resulting in an empty network if the selected feature renders an empty set of concepts (e.g., a user has not made any edits yet).

The author specific concept networks can be used to analyze a variety of different aspects related to user behavior, such as the role of a specific user (i.e. generalists vs. specialists, see Figure 3) during the engineering process or concepts, topics and areas of interest of specific users.

In addition to these features and visualizations, we have also implemented a heat-map (see Figure 4), which allows users to visually monitor and track activity within the ontology. The heat-map can be used in all concept networks (including all user specific concept networks) combined with every feature listed in Table 1.

The *Administrative Personnel* can use the concept network visualization to track activity and progress as well as to identify domain specialists. The *Engineering Staff* can use the tool to identify parts of the ontology that are (or are not) very active to adapt the engineering process or the underlying knowledge representation. *Ontology Viewers* can use the visual representation to explore the complexity of the ontology and to identify areas of community interest according to different pragmatic features (e.g. number of edits).

Author Network Visualization

In addition to the user specific concept network, *PragmatiX* provides a visualization of all authors, displaying the extent of collaboration (see Figure 2) they engaged in during the engineering process. The following two features can be selected:

1. *Commonly edited categories (collaboration*) shows a network of authors, connected by weighted edges according to the number of commonly changed or commented on concepts. The node size represents the total number of changes performed by each author.
2. *Overrides* shows a network of authors, connected by weighted edges according to the number of changes by one author that were overridden by another author. The node size represents the fraction of all changes performed by the author that were overridden by other authors.

NIH-PA Author Manuscript

NIH-PA Author Manuscript

NIH-PA Author Manuscript

The author network supports *Administrative Personnel* to measure if and to what extent authors engage in collaboration and perform overrides. Additionally it can help members of the *Engineering Staff* to explicitly identify the "*importance*" of a user (e.g. for the collaboration graph, according to the connectivity and weights of the edges, indicating very active and collaborative users). *Ontology Viewers* will use the author network to explore the complexity of social interactions and might be interested to compare the extent of collaboration with the amount of performed overrides to see whether the project is led in a more or less democratic way and who is responsible for keeping the order and can be contacted in case problems arise.

Property Network Visualization

In collaborative ontology engineering projects, it is interesting to study the pragmatic relationships between different properties, to show – for example - what properties are edited first, or in what sequence properties are edited (where a property refers to a property of the ontology, i.e. a data type, object or annotation property). This could allow for identifying patterns of property editing behavior, which could have implications for the design of more effective user interfaces. To that end, we calculated and visualized the property network. Each node corresponds to a property and every edge represents the number of changes on a property followed by a change on a different property.

The relationships between properties could be of great interest for the *Engineering Staff* and explicitly for *Knowledge Engineers* to enhance and adapt their knowledge representation and tools to better fit the natural working process of its users by closely grouping properties that exhibit highly weighted edges.

Network Visualization Implementation Details

The nodes, which represent either concepts or users, and edges, which represent either ontological relations or collaboration and overrides, in the network visualizations are weighted according to a set of independent features (see Table 1 for more details) and are visualized by adjusting the size and/or color of the nodes and edges. Currently, *PragmatiX* features the following network visualization layouts, both generated using Graphviz (Ellson et al., 2001):

1. *twopi* (radial)
2. *sfdp* (multi-scale force-directed "spring model")

The radial layout allows for a clear visualization of ontological or hierarchical structures similar to that of taxonomies or trees. The force-directed layout, on the other hand, is better suited to visualize highly interlinked ontological structures and networks. Due to the fact that all layouts are pre-calculated (i.e. the x- and y-positions for each concept are stored in the database for each layout), *PragmatiX* can be easily extended to support any given 2-dimensional layout algorithm and does not necessarily depend on Graphviz.

To navigate the graphical representations users can either use the arrow-keys on their keyboard combined with the graphical user interface that allows for easy zooming and jumping back to the center of the network, or they can explore the network by common drag-n-drop principles and adjust the zoom level by using the mouse wheel.

For reasons of usability and to avoid visual cluttering, *PragmatiX* only displays a specific fraction of nodes for large data sets, rather than the whole network at once. This is aimed to enable users in identifying and exploring top (i.e. the most interesting) concepts regarding specific features and attributes rather than analyzing the layout of an imported data set in general.

NIH-PA Author Manuscript

NIH-PA Author Manuscript

NIH-PA Author Manuscript

To that end, *PragmatiX* knows the coordinates of the user view's bounding box and selects the corresponding part of the network to display. To determine which nodes are displayed to avoid visual cluttering in large data sets, we have implemented an intelligent filtering algorithm, which divides the bounding box (or field of view) into 10x10 raster boxes where each box displays the node with the highest weight of the currently selected feature within its boundaries. To avoid disconnecting components that are physically connected, all nodes from any selected node and all edges from any selected node to the root node are displayed as well if available, thus forcing the network to stay connected.

Dashboard

The *dashboard* (see Figure 5) was created to provide something similar to an “overview” page that provides overall and generalized statistics about the whole collaborative ontology-engineering project.

Figure 5a shows the changes and notes distribution over time as a line chart and visualizes the aggregated amount of performed notes and changes in the ontology over time. The distribution of changes across users (Figure 5b) is represented as a pie chart. The basic statistics table (Figure 5c) is a textual representation to quantify the size of the ontology, its users and their performed changes and annotations. Additionally, the category display status statistics (Figure 5d), provides additional information about the amount of concepts (and their average number of changes) with the corresponding assigned display states.

The Community Statistics pie charts (Figure 5d) provide information about the percentage of changes performed by each user group on their corresponding concepts. For example, as can be seen in Figure 5, the corresponding assigned primary community performed 25.99% of all changes across all concepts.

The *dashboard* was specifically designed to fit the requirements of the *Administrative Personnel* to provide a quick overview of the current progress of the ontology. Additionally it supports the *Engineering Staff* to identify the distribution of edits across time and users. *Ontology Viewers* might be interested in parts of the *dashboard*, depending on their personal motivation towards *PragmatiX* and the imported data set.

Community Views

If an ontology exhibits different Communities of Interest (or user groups), *PragmatiX* provides “smaller *dashboards*” for each community, called “*community views*”. They are analogously designed to the dashboard (see Figure 5) and provide the same graphical visualizations, displaying only relevant data for each community.

In addition to the dashboard, each community view displays textual information of the amount of changes and notes performed on the concepts, which are assigned to the community as well as the total number of changes and notes performed on these concepts by all authors in the ontology.

Community views are specifically designed to meet the requirements of *Project Managers* and their assigned areas of the ontology. Similar to the dashboard, both *Administrative Personnel* and *Ontology Viewers* might be interested in specific community views, depending on their current motivation or tasks. In the case of the ICD-11, a community of interest is usually referred to as a Topic Advisory Group (TAG).

NIH-PA Author Manuscript

NIH-PA Author Manuscript

NIH-PA Author Manuscript

Category and Author Views

In addition to the network visualizations, *PragmatiX* provides overviews and very detailed statistics views for all authors and categories in general and every single author and category in detail. In these overviews, we rank all concepts and authors according to our implemented features (see Table 1 for a list of features). This allows users to quickly identify the top (and worst) concepts (see Figure 6) or authors for every feature without having to browse the graphical network visualizations.

The detailed concept and author statistics views can be reached by either clicking on a node in the corresponding network visualizations or by following the links (as displayed in Figure 6) on the ranked overviews. These links are represented by the title of each concept and in the case of ICTM, which provides multiple languages for each concept; all available title-translations for the corresponding concepts are displayed in the listings of Figure 6.

The detailed concept statistics views (see Figure 7) provide further information about the parents and children of a concept, the change and note history of a concept, the group dynamics (e.g. Who contributed what amount of edits or notes when?) and a table that lists all feature-values used in the concept network visualization. Due to reasons of space, the table containing the pre-calculated feature-values has been omitted in Figure 7.

The detailed author statistics views are similar to the detailed concept statistics views and provide information about the amount of changes and notes contributed by a specific user, the communities a user is member of, concept-recommendations (Walk et al., 2012) the user might be interested to change as well as co-authors and overrides performed on a user.

The detailed statistics views for each concept and every author provide useful information for all members of the *Engineering Staff*. Project Managers and Knowledge Engineers can reconstruct the change history of a concept and identify the corresponding most influential/active users using the detailed statistics concept views. Knowledge Providers can, for example, use their own detailed statistics author view to receive suggestions for concepts to edit. Both concept and author overviews are used by *Administrative Personnel* and the *Engineering Staff* to quickly identify users or concepts of interest, according to the implemented and ranked by features.

EVALUATION

PragmatiX was evaluated in May 2012 using the Heuristic Evaluation (HE) method. Three ontology engineering experts, who are all experienced in the task of engineering and gardening ontologies, thus represent a fraction of the actual target group of our tool, acted as evaluators and explored, tested and investigated the interface in sessions from 60 to 120 minutes. All three evaluators, without any direct request, assumed (multiple) different user roles during the evaluation task and included these perspectives in the provided feedback. The evaluation uncovered a total of 27 usability issues, ranging from rather simple problems, such as a misleading icon to display a legend while browsing a network visualization, to more serious usability issues, such as a confusing name for the concept of TAGs (Topic Advisory Groups, introduced by the WHO for ICD-11) which are better described as Communities of Interest. All 27 issues were classified according to a modified version of Nielsen's 10 Heuristics (Nielsen, 1994) called the “Andrews General Usability Heuristics”, which are more concise and include small explanations or examples as clarification for each heuristic, which aid the evaluators in classifying identified problems during their evaluation task. We make the full HE Report available for download as Walk & Andrews (2012). A short excerpt of identified usability issues, manually filtered by significance and ranked according to their severity, is listed in Table 2.

NIH-PA Author Manuscript

NIH-PA Author Manuscript

NIH-PA Author Manuscript

All three evaluators, who have expertise and experience in developing and working with ontologies stated, that they were confused about the initial concept network visualization they were presented immediately after login. According to the feedback gathered during the HE, the confusion mainly arose as the evaluators never specified any features or relationships prior to the login and could not link the displayed information with the visualization. Additionally all information regarding the chosen data set, the selected features and the explanations of these are hidden within the interface of *PragmatiX*. The evaluators mentioned, that they are missing explicit information about the currently displayed visualization, which can help them to better understand what they are currently looking at. One evaluator specifically stated, that she is missing information on which nodes are currently displayed.

This leads to another very interesting observation. The evaluators had problems to really understand and interpret the different features. It was not immediately clear why it could be of interest to explore and visualize the number of changes performed on each concept or their number of distinct authors.

According to the evaluators this is mainly due to the lack of descriptions for the implemented features, their very unspecific presentation (i.e. describing the drop-down box to select the nodesize features with “Feature”, rather than “Feature to define node-size:”) combined with an extensive usage of network theory vocabulary, which ontology experts might not be familiar with. One evaluator suggested, that we should use descriptive text-snippets rather than the actual name of the implemented measures.

This also correlates with the problems the evaluators faced, when trying to interpret the meaning of the visualizations after selecting different features. Not only was it unclear to the evaluators, what property of which element is influenced by selecting different features, but also how to make sense of the visualizations of these features.

In a few cases bigger nodes concealed smaller nodes or the displayed additional information when hovering a node was unreadable due to overlapping neighboring nodes, which actually amplified the problem of interpretation and is a direct result of pre-calculated graph layouts, which do not accommodate different node sizes.

As a possible solution to the problem of interpreting the visualizations, we were asked to provide additional textual information on the meaning of the currently selected feature and maybe some smaller examples to explain how to interpret the measure.

The majority of all uncovered problems are related to providing more and additional information about the implementation and meaning of various attributes of *PragmatiX*. All three evaluators mentioned the high utility of the implemented heat-map, the dashboard and community views.

It is noteworthy that, in addition to usability issues and positive findings, two evaluators also explicitly tried to locate specific features, which were not yet part of the system. This led to the introduction of a third category of findings called “*feature requests*”. One of these feature requests was, for example, the implementation of a timeline graph that would support browsing the state of the collaboratively engineered ontology and its engineering process at different points in time.

Additionally, it is important to note that *PragmatiX* is the first version of the tool, which was properly evaluated, thus we only have informal information about issues and complaints regarding its predecessor *iCAT Analytics*, which are all related to missing features. Nonetheless, the biggest complaints mentioned by users of *iCAT Analytics* were the

NIH-PA Author Manuscript NIH-PA Author Manuscript NIH-PA Author Manuscript

limitation of the tool to only support one data set per instance and the lack of group statistics, which were especially interesting to members of WHO.

Further improvements and formative evaluations are anticipated, and at some point in the future a summative study with end users from WHO is planned. However, the involvement of end users from WHO requires extensive planning and coordination, which is why further evaluations and the summative study are subject of future work.

In general, the visualization of very large networks is a very hard task, especially when it is desired to use the output for explorative analysis. The performed HE suggests that the currently implemented interface is useful for explorative analysis by browsing the visualizations, however new approaches and support have to be implemented for helping users in interpreting the visualized results.

In general, the feedback received during the HE was very valuable as it helped to uncover multiple design flaws that potentially confuse our target user-groups and aggravate dealing not only with *PragmatiX* but also collaboratively engineered ontologies in general. On the other hand, specific advice was gathered to derive potential approaches that can help to solve the identified usability flaws (i.e. to further explain the colors in all charts or to better describe the measures throughout the tool instead of only providing the name of the measure itself). However, it should be noted that an HE does not guarantee to identify all usability issues and is not designed to provide solutions to positively identified issues.

DISCUSSION

PragmatiX aims to visualize pragmatic aspects of the creation processes behind collaboratively engineered ontologies. We argue that the three implemented network visualization views are useful for a variety of exploratory tasks. For example, the concept network visualization can be used to monitor progress, identify generalists and specialists, to detect areas and concepts of high (and recent/past) activity. The author network visualization can be used to measure if and to what extent collaboration and overrides exist in the project and which authors collaborate with, or override other authors the most. The properties network visualization can provide insights into the creation process, which in turn can help to enhance the engineering tools for example by grouping properties that are frequently and successively changed. Our heuristic evaluation has demonstrated that – in principle – the *PragmatiX* tool can serve these purposes. It found that the majority of problems assigned to *PragmatiX* are related to insufficient descriptive textual information of implemented features. As a result, evaluators confused or misinterpreted vocabulary used in *PragmatiX*, which is mostly taken from the domain of network analysis. *PragmatiX* analyzes and visualizes the edit-, and contribution-behavior of all users that have contributed to a project and as a result are also named in the inspected change-logs, which can be a great benefit for management. Nonetheless, privacy poses an open issue and has to be addressed by the corresponding Ontology Administrators. One possible solution to this problem could be to obscure the change-logs by replacing names with consistent acronyms prior to importing them into *PragmatiX*. In the long run, identifying more profound, automatic and secure approaches to protect contributors privacy poses a very important subject of future work.

CONCLUSIONS

In this application paper, we have presented *PragmatiX* - a tool for visualizing the construction processes behind collaborative ontology engineering projects. Our main motivation for the development of this new tool was (i) an interest in making the otherwise hidden social processes and dynamics behind collaborative ontology engineering more

NIH-PA Author Manuscript NIH-PA Author Manuscript NIH-PA Author Manuscript

visible and amenable to analysis and (ii) a lack of currently available visualization tools for that purpose. We have presented and preliminarily evaluated the main functionality and features of *PragmatiX*, and we have demonstrated its general applicability by using it for visualizing pragmatic aspects of five collaborative ontology-engineering projects. However due to limitations in space, only two projects are presented in this paper.

We aimed to demonstrate that *PragmatiX* is a promising tool to visualize and analyze the pragmatic processes behind large collaborative ontology engineering projects. Future work on *PragmatiX* will likely focus on usability issues and feature requests gathered during the heuristic evaluation. We also anticipate including additional graph or network layouts that potentially provide additional insights into the social fabric, which would help to identify cliques or other groups of collaborators. Finally, we hope that this work sparks a new line of research on visualization tools for analyzing the processes behind collaborative ontology engineering projects.

REFERENCES

Alani, H. TGVizTab: An Ontology Visualisation Extension for Protégé.. Knowledge capture 03 - workshop on visualizing information in knowledge engineering; Sanibel Island, FL: ACM. 2003. p. 2-7.

Auer, S.; Dietzold, S.; Riechert, T. OntoWiki - A Tool for Social, Semantic Collaboration.. In: Cruz, IF., et al., editors. The Semantic Web - ISWC 2006, 5th International Semantic Web Conference, ISWC 2006, Athens, GA, USA, November 5-9, 2006, Proceedings; Berlin / Heidelberg: Springer. 2006. p. 736-749.

Bao, J.; Honavar, V. Collaborative ontology building with wiki@nt - a multi-agent based ontology building environment.. Proceedings of the 3rd International Workshop on Evaluation of Ontology-Based Tools; Oct. 2004 p. 1-10.

Barnum, CM. Usability *Testing Essentials: Ready, Set... Test!*. Morgan Kaufmann Publishers; Burlington, MA: 2010.

Bosca, A.; Bonino, D.; Pellegrino, P. Bouquet, P.; Tummarello, G., editors. Ontosphere: More than a 3D Ontology Visualization Tool.. Swap. 2005.

Brank, J.; Grobelnik, M.; Mladenic, D. A survey of ontology evaluation techniques.. Proceedings of the Conference on Data Mining and Data Warehouses; Lubljana, Slovenia. 2005. p. 166-170.

Cristani M, Cuel R. A survey on ontology creation methodologies. International Journal on Semantic Web & Information Systems. 2005; 1(2):49–69.

Eklund, PW.; Roberts, N.; Green, SP. The first international symposium on cyberworlds. IEEE press; 2002. Ontorama: Browsing an RDF ontology using a hyperbolic-like browser.; p. 405-411.

Ellson, J.; Gansner, ER.; Koutsofios, E.; North, SC.; Woodhull, G. Graphviz - open source graph drawing tools.; Graph drawing. 2001. p. 483-484.

Falconer, S. [October 17, 2012] OntoGraf.. Protégé Wiki. Apr 12. 2010 from http://protegewiki.stanford.edu/wiki/OntoGraf

Falconer, SM.; Tudorache, T.; Noy, NF. An Analysis of Collaborative Patterns in Large-Scale Ontology Development Projects.. In: Musen, MA.; Corcho, Oscar, editors. Proceedings of the sixth international conference on knowledge capture; New York, NY: ACM. 2011. p. 25-32.

Golbeck J, Fragoso G, Hartel FW, Hendler JA, Oberthaler J, Parsia B. The National Cancer Institute's Thsaurus and Ontology. Journal of Web Semantics. 2003; 1(1):75–80.

Hagberg, AA.; Schult, DA.; Swart, PJ. Exploring network structure, dynamics, and function using networkx.. In: Varoquaux, G.; Vaught, T.; Millman, J., editors. Proceedings of the 7th python in science conference; Pasadena, CA USA. 2008. p. 11-15.

Harris MA, Clark J, Ireland A, Lomax J, Ashburner M, Foulger R, Eilbeck K, Lewis S, Marshall B, Mungall C, et al. The Gene Ontology (GO) database and informatics resource. Nucleic Acids Research. 2004; 32(Database Issue):258–261.

Horridge, M. [June 02, 2012] OWLViz – A visualization plugin for the Protégé OWL Plugin. 2012. from http://www.co-ode.org/downloads/owlviz/OWLVizGuide.pdf

NIH-PA Author Manuscript

NIH-PA Author Manuscript

NIH-PA Author Manuscript

Lanzenberger, M.; Sampson, J. Alviz - A Tool for Visual Ontology Alignment.. Proceedings of the conference on Information Visualization; Washington, DC, USA:IEEE Computer Society. 2006. p. 430-440.

Maedche, A.; Staab, S. Semi-automatic engineering of ontologies from text.. Proceedings of the 12th Internal Conference on Software and Knowledge Engineering; Chicago: KSI. 2000.

Nielsen, J. Enhancing the Explanatory Power of Usability Heuristics.. Proceedings of the SIGCHI Conference on Human Factors in Computing Systems: Celebrating Interdependence; New York, NY, USA:ACM. 1994. p. 152-158.

Nielsen, J.; Mack, RL. Usability Inspection Methods. John Wiley & Sons; New York, NY: 1994.

Noy N, Chugh A, Liu W, Musen ME. A framework for ontology evolution in collaborative environments. The Semantic Web - ISWC 2006. 2006:544–558.

Noy, N.; McGuinness, D., et al. Ontology development 101: A guide to creating your first ontology.. Stanford knowledge systems laboratory technical report KSL-01-05 and Stanford medical informatics technical report SMI-2001-0880. 2001. Retrieved from http://www.ksl.stanford.edu/people/dlm/papers/ontology-tutorial-noy-mcguinness-abstrac

Noy NF, Tudorache T. Collaborative ontology development on the (semantic) web. Nature Biotechnology. 2008

Pöschko, J.; Strohmaier, M.; Tudorache, T.; Noy, NF.; Musen, MA. Pragmatic analysis of crowd-based knowledge production systems with iCAT Analytics: Visualizing changes to the ICD-11 ontology.. Proceedings of the AAAI Spring Symposium 2012: Wisdom of the Crowd; 2012.

Rubin, J.; Chisnell, D.; Spool, J. Handbook of Usability Testing: How to Plan, Design, and Conduct Effective Tests. 2. ed.. John Wiley & Sons, Inc.; Indianapolis, IN: 2008.

Schreiber, G.; Akkermans, H.; Anjewierden, A.; de Hoog, R.; Shadbolt, N.; Van de Velde, W.; Wielinga, B. Knowledge Engineering and Management: The CommonKADS Methodology. MIT Press; Cambridge, MA: 2000.

Singh, G.; Prabhakar, T.; Chatterjee, J.; Patil, V.; Ninomiya, S., et al. Ontoviz: Visualizing Ontologies and Thesauri using Layout Algorithms.. Afita 2006: The fifth international conference of the asian federation for information technology in agriculture, jn tata auditorium, indian institute of science campus; Bangalore, India. 9-11 November, 2006; 2006. p. 709-719.

Spyns P, Meersman R, Jarrar M. Data Modelling versus Ontology Engineering. SIGMOD Rec. Dec; 2002 31(4):12–17.

Storey, M-A.; Best, C.; Michaud, J.; Rayside, D.; Litoiu, M.; Musen, M. Chi '02 Extended Abstracts on Human Factors in Computing Systems. ACM; New York, NY, USA: 2002a. Shrimp Views: An Interactive Environment for Information Visualization and Navigation.; p. 520-521.

Storey, M-A.; Noy, NF.; Musen, M.; Best, C.; Fergerson, R.; Ernst, N. Jambalaya: an interactive environment for exploring ontologies.. Proceedings of the 7th International Conference on Intelligent User Interfaces; New York, NY, USA: ACM. 2002b. p. 239-239.

Sure, Y.; Erdmann, M.; Angele, J.; Staab, S.; Studer, R.; Wenke, D. OntoEdit: Collaborative Ontology Development for the Semantic Web.. In: Horrocks, I.; Hendler, JA., editors. International Semantic Web Conference; Springer. 2002. p. 221-235.

Tudorache, T.; Falconer, S.; Noy, NF.; Nyulas, C.; Üstün, TB.; Storey, M-A.; Musen, MA. Ontology development for the masses: Creating ICD-11 in WebProtégé. Proceedings of the 17th International Conference on Knowledge Engineering and Management by the Masses; Berlin, Heidelberg: Springer-Verlag. 2010. p. 74-89.

Tudorache, T.; Falconer, SM.; Nyulas, C.; Noy, NF.; Musen, MA. Will Semantic Web Technologies Work for theDevelopment of ICD-11?. In: Patel-Schneider, PF., et al., editors. International Semantic Web Conference (2); Springer. 2010. p. 257-272.

Tudorache, T.; Noy, NF.; Tu, S.; Musen, MA. Supporting Collaborative Ontology Development in Protégé.. Proceedings of the 7th International Semantic Web Conference 2008; Karlsruhe, Germany: Springer. Oct. 2008 p. 17-32.

Tudorache T, Nyulas C, Noy NF, Musen MA. WebProtégé: A Distributed Ontology Editor and Knowledge Acquisition Tool for the Web. Semantic Web Journal. 2011:11–165.

Walk, S.; Andrews, K. [July 6, 2012] Heuristic Evaluation Report. 2012. from http://heuristiceval.simonwalk.at/he.html

NIH-PA Author Manuscript

NIH-PA Author Manuscript

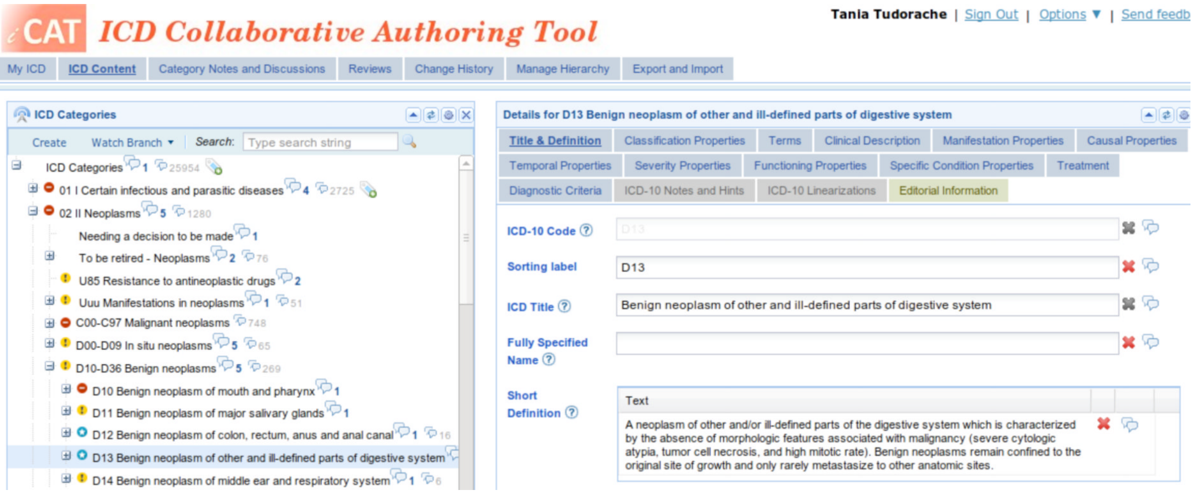
NIH-PA Author Manuscript

Walk, S.; Strohmaier, M.; Tudorache, T.; Noy, NF.; Nyulas, C.; Musen, MA. Recommending Concepts to Experts: An Exploration of Recommender Techniques for Collaborative Ontology Engineering Platforms in the Biomedical Domain.. Proceedings of the 3rd International Conference on Biomedical Ontology (ICBO 2012); Graz, Austria. 2012; 2012.

NIH-PA Author Manuscript

NIH-PA Author Manuscript

NIH-PA Author Manuscript


iCAT ICD Collaborative Authoring Tool
Tania Tudorache | Sign Out | Options ▼ | Send feedb
My ICD
ICD Content
Category Notes and Discussions
Reviews
Change History
Manage Hierarchy
Export and Import
ICD Categories
Create
Watch Branch ▾
Search: Type search string
ICD Categories 1 25954
01 I Certain infectious and parasitic diseases 4 2725
02 II Neoplasms 5 1280
Needing a decision to be made 1
To be retired - Neoplasms 2 76
U85 Resistance to antineoplastic drugs 2
Uuu Manifestations in neoplasms 1 51
C00-C97 Malignant neoplasms 748
D00-D09 In situ neoplasms 5 65
D10-D36 Benign neoplasms 5 269
D10 Benign neoplasm of mouth and pharynx 1
D11 Benign neoplasm of major salivary glands 1
D12 Benign neoplasm of colon, rectum, anus and anal canal 1 16
D13 Benign neoplasm of other and ill-defined parts of digestive system
D14 Benign neoplasm of middle ear and respiratory system 1 6
Details for D13 Benign neoplasm of other and ill-defined parts of digestive system
Title & Definition
Classification Properties
Terms
Clinical Description
Manifestation Properties
Causal Properties
Temporal Properties
Severity Properties
Functioning Properties
Specific Condition Properties
Treatment
Diagnostic Criteria
ICD-10 Notes and Hints
ICD-10 Linearizations
Editorial Information

ICD-10 Code: D13

Sorting label: D13

ICD Title: Benign neoplasm of other and ill-defined parts of digestive system

Fully Specified Name:

Short Definition: Text

A neoplasm of other and/or ill-defined parts of the digestive system which is characterized by the absence of morphologic features associated with malignancy (severe cytologic atypia, tumor cell necrosis, and high mitotic rate). Benign neoplasms remain confined to the original site of growth and only rarely metastasize to other anatomic sites.

Figure 1.

Exemplary Screenshot of the ICD Collaborative Authoring Tool (iCAT) that is used to create and maintain ICD-11.

NIH-PA Author Manuscript

NIH-PA Author Manuscript

NIH-PA Author Manuscript

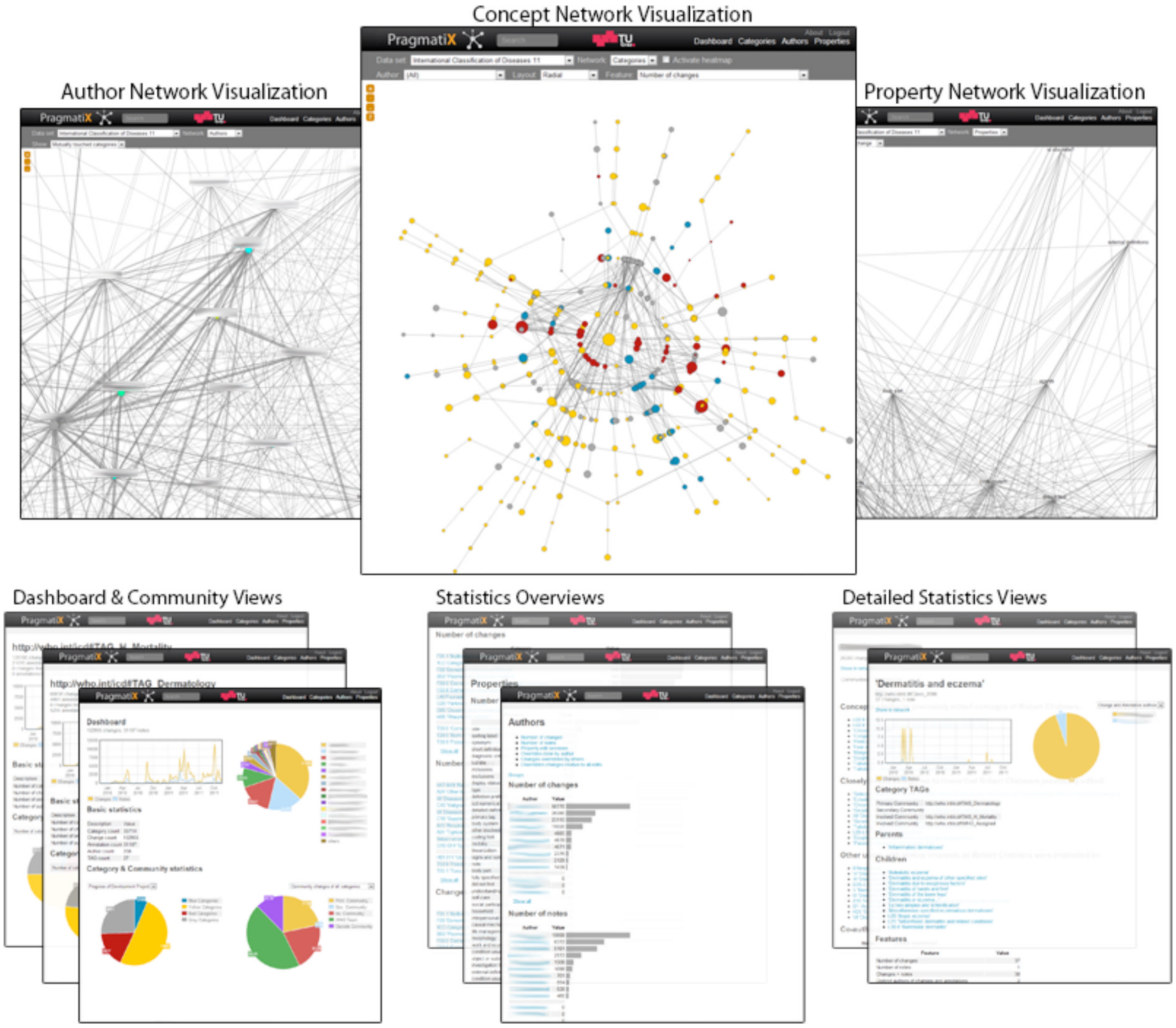

Concept Network Visualization
Author Network Visualization
Property Network Visualization
PragmatiX
Dashboard Categories Authors Properties
Data set: International Classification of Diseases 11
Network: Categories
Activate heatmap
Author: (All)
Layout: Radial
Feature: Number of changes
Dashboard & Community Views
Statistics Overviews
Detailed Statistics Views
Dashboard
Basic statistics
Category & Community statistics
Authors
Number of changes
Number of notes
'Dermatitis and eczema'
Category TAGs
Parents
Children
Features

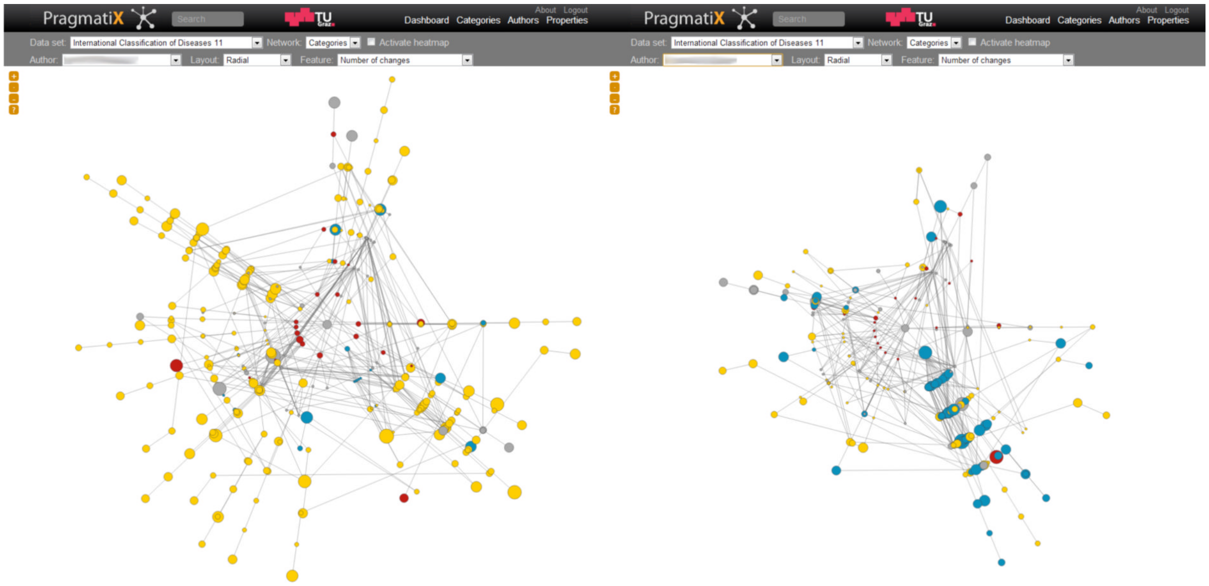
Figure 2.

An overview of all statistics views and visualizations provided by PragmatiX. The tool features three network visualizations (Author, Concept and Property Network), different types of dashboards and overviews as well as very detailed statistics views. The data displayed in this figure is from the ICDG11 data set. The nodes in the network views either represent an author, a concept or a property. While edges in the concept network visualization depict isGa relationships, they vary in the author and properties network visualization according to the currently selected features.

NIH-PA Author Manuscript

NIH-PA Author Manuscript

NIH-PA Author Manuscript



Data set: International Classification of Diseases 11
Network: Categories
□ Activate heatmap
Author: [redacted]
Layout: Radial
Feature: Number of changes

Data set: International Classification of Diseases 11
Network: Categories
□ Activate heatmap
Author: [redacted]
Layout: Radial
Feature: Number of changes

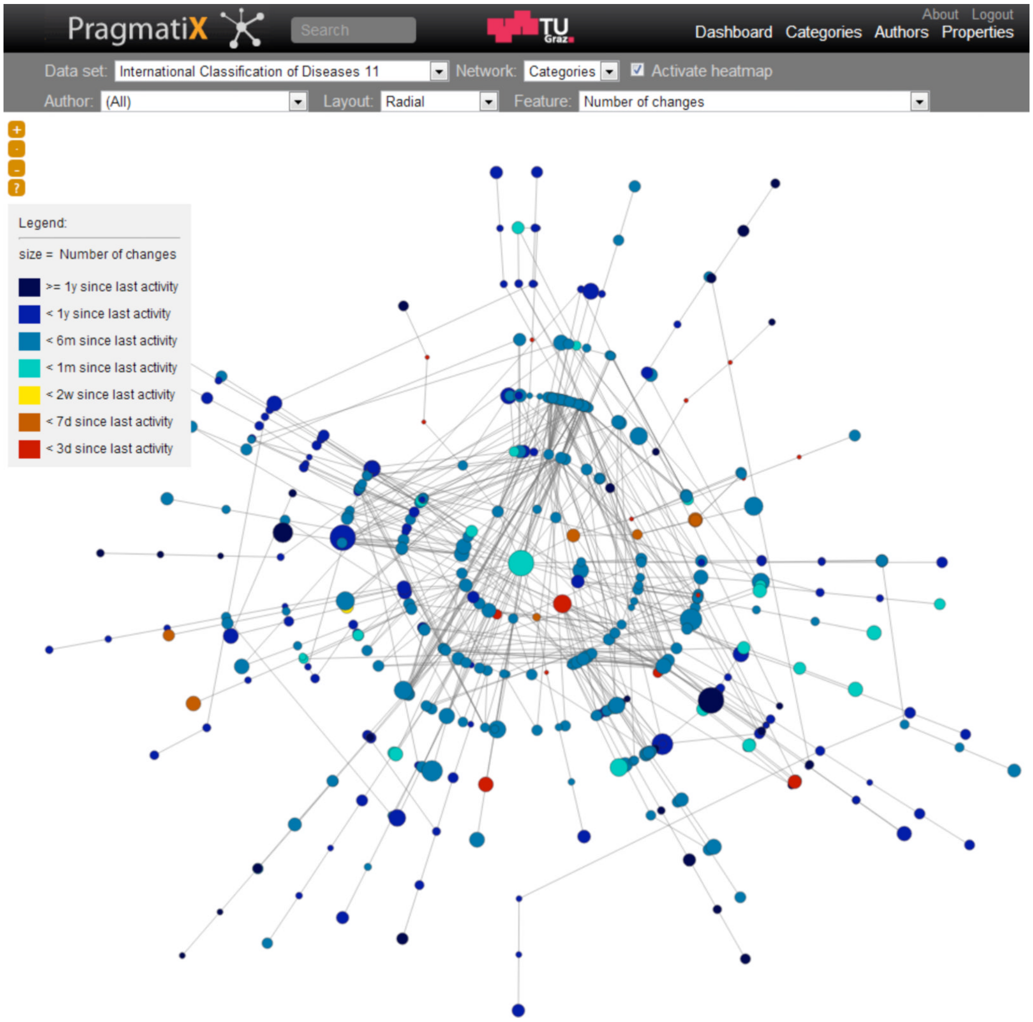
Figure 3.

Two author specific concept network visualizations taken from the ICD-11 data set with the selected feature “Number of changes”. The left author-specific concept network indicates a user who seems to be more of a generalist (editing a vast variety of concepts in the ontology) while the right concept network indicates a user who seems to be more of a specialist (editing a more narrow area of the ontology).

NIH-PA Author Manuscript

NIH-PA Author Manuscript

NIH-PA Author Manuscript



Data set: International Classification of Diseases 11
Network: Categories
☑ Activate heatmap
Author: (All)
Layout: Radial
Feature: Number of changes

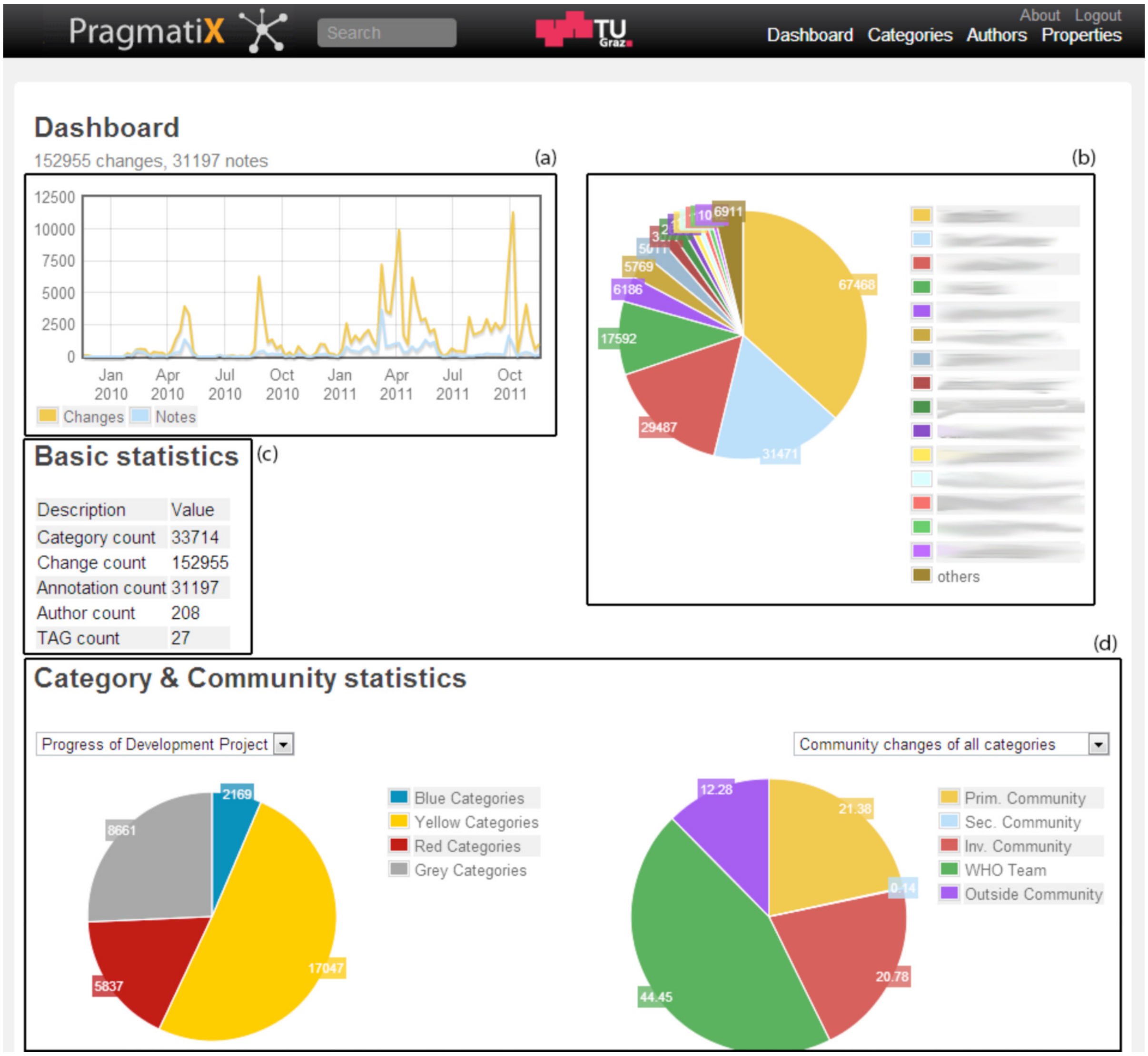
Figure 4.

PragmatiX showing the concept network visualization of ICD-11 as a heat-map and the selected feature “Number of changes”. The warmer the node colors the more recent the last edit has been performed.

NIH-PA Author Manuscript

NIH-PA Author Manuscript

NIH-PA Author Manuscript



Basic statistics

Description	Value
Category count	33714
Change count	152955
Annotation count	31197
Author count	208
TAG count	27

Category & Community statistics

Progress of Development Project

Community changes of all categories

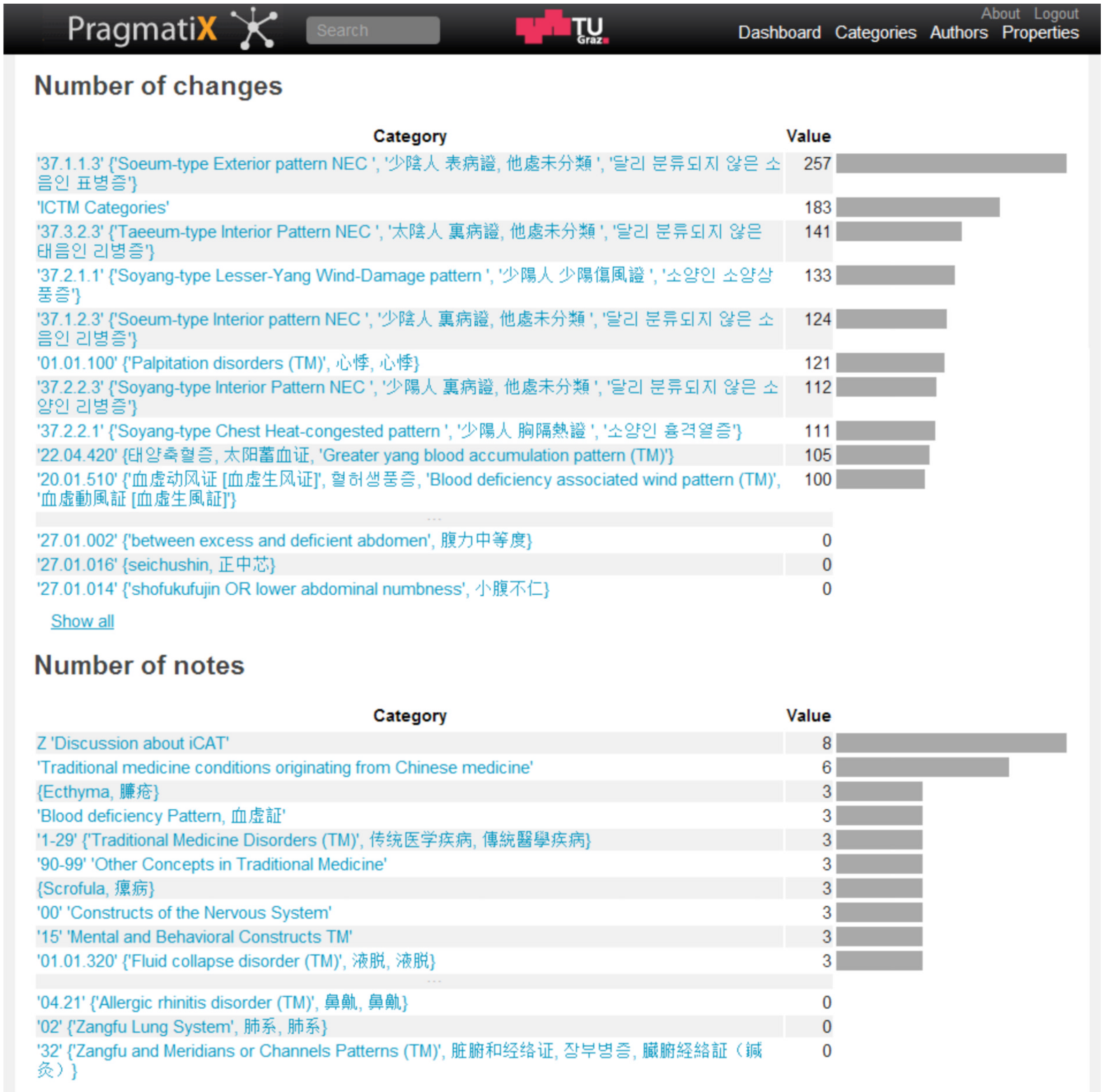
Figure 5.

A screenshot of the dashboard, providing analytical information for the complete ICD-11 engineering process, divided into four elements. Figure 5a shows the total amount of notes and changes performed over time. Figure 5b shows the amount of changes performed by the Top 15 contributors. The Basic Statistics table (Figure 5c) can be used to quantify the size of the ontology, its users and their performed changes and annotations. The Concept and Community Statistics (Figure 5d) provide additional details about the distribution of changes across assigned communities and the overall progress of the project.

NIH-PA Author Manuscript

NIH-PA Author Manuscript

NIH-PA Author Manuscript



Number of changes

Category	Value
'37.1.1.3' {'Soeum-type Exterior pattern NEC ', '少陰人 表病證, 他處未分類 ', '달리 분류되지 않은 소음인 표병증'}	257
'ICTM Categories'	183
'37.3.2.3' {'Taeeum-type Interior Pattern NEC ', '太陰人 裏病證, 他處未分類 ', '달리 분류되지 않은 태음인 리병증'}	141
'37.2.1.1' {'Soyang-type Lesser-Yang Wind-Damage pattern ', '少陽人 少陽傷風證 ', '소양인 소양상풍증'}	133
'37.1.2.3' {'Soeum-type Interior pattern NEC ', '少陰人 裏病證, 他處未分類 ', '달리 분류되지 않은 소음인 리병증'}	124
'01.01.100' {'Palpitation disorders (TM)', 心悸, 心悸}	121
'37.2.2.3' {'Soyang-type Interior Pattern NEC ', '少陽人 裏病證, 他處未分類 ', '달리 분류되지 않은 소양인 리병증'}	112
'37.2.2.1' {'Soyang-type Chest Heat-congested pattern ', '少陽人 胸隔熱證 ', '소양인 흉격열증'}	111
'22.04.420' {태양축혈증, 太阳蓄血证, 'Greater yang blood accumulation pattern (TM)'}	105
'20.01.510' {'血虚动风证 [血虚生风证]', 혈허생풍증, 'Blood deficiency associated wind pattern (TM)', '血虛動風証 [血虛生風証]'}	100
...	
'27.01.002' {'between excess and deficient abdomen', 腹力中等度}	0
'27.01.016' {seichushin, 正中芯}	0
'27.01.014' {'shofukufujin OR lower abdominal numbness', 小腹不仁}	0

Show all

Number of notes

Category	Value
Z 'Discussion about iCAT'	8
'Traditional medicine conditions originating from Chinese medicine'	6
{Ecthyma, 臁疮}	3
'Blood deficiency Pattern, 血虛証'	3
'1-29' {'Traditional Medicine Disorders (TM)', 传统医学疾病, 傳統醫學疾病}	3
'90-99' 'Other Concepts in Traditional Medicine'	3
{Scrofula, 瘰疬}	3
'00' 'Constructs of the Nervous System'	3
'15' 'Mental and Behavioral Constructs TM'	3
'01.01.320' {'Fluid collapse disorder (TM)', 液脱, 液脫}	3
...	
'04.21' {'Allergic rhinitis disorder (TM)', 鼻鼽, 鼻鼽}	0
'02' {'Zangfu Lung System', 肺系, 肺系}	0
'32' {'Zangfu and Meridians or Channels Patterns (TM)', 脏腑和经络证, 장부병증, 臟腑経絡証（鍼灸）}	0

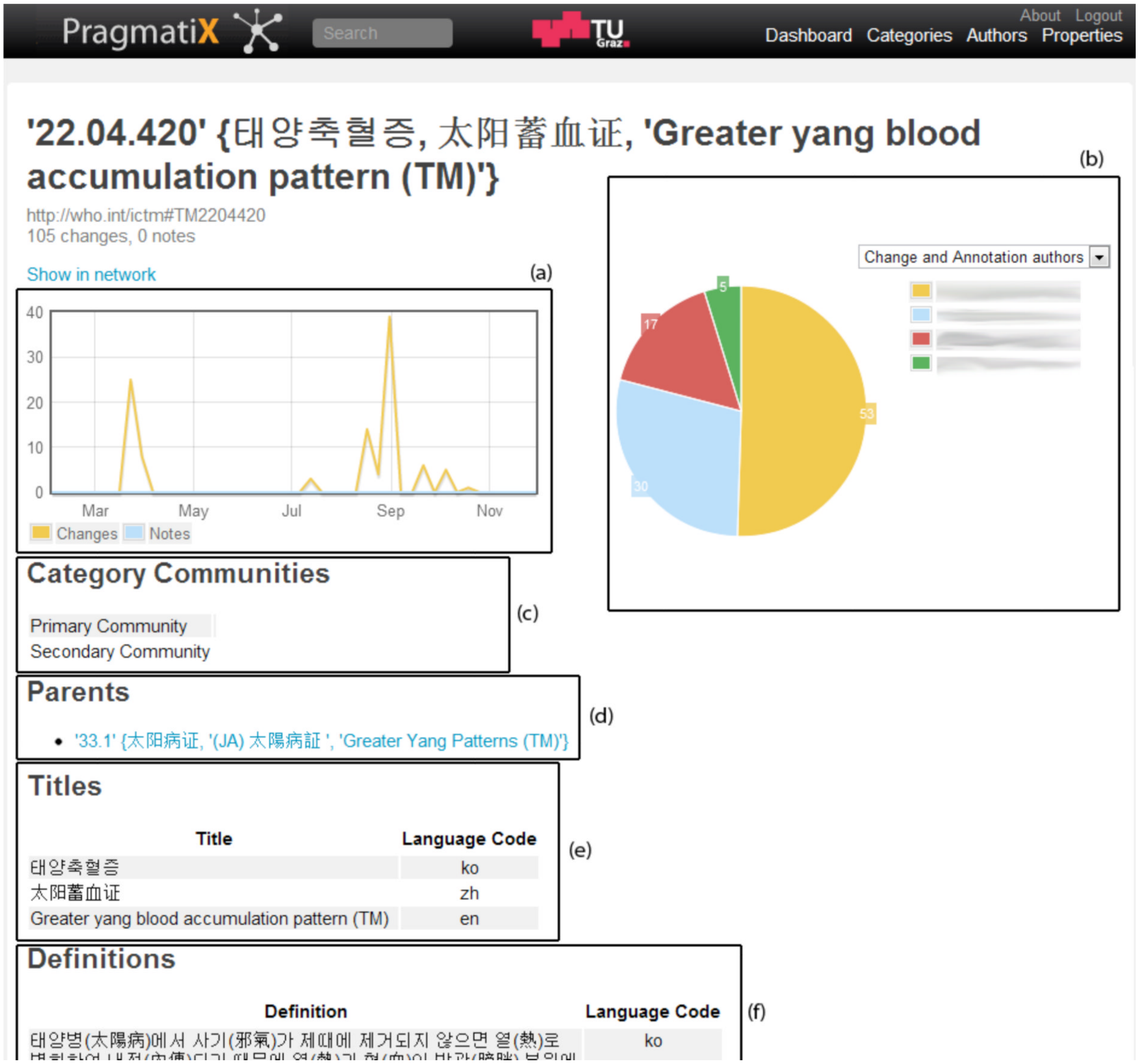
Figure 6.

An excerpt of the concept overview for ICTM showing the Top 10 (and worst 3) concepts according to the features “Number of Changes” and “Number of notes” as clickable links. As ICTM provides multiple languages for each concept title and definition, all available titles for the corresponding concepts are displayed in the listings of Figure 6.

NIH-PA Author Manuscript

NIH-PA Author Manuscript

NIH-PA Author Manuscript



'22.04.420' {태양축혈증, 太阳蓄血证, 'Greater yang blood accumulation pattern (TM)'}

http://who.int/ictm#TM2204420
105 changes, 0 notes

Show in network

Category Communities

Primary Community
Secondary Community

Parents

- '33.1' {太阳病证, '(JA) 太陽病証 ', 'Greater Yang Patterns (TM)'}

Titles

Title	Language Code
태양축혈증	ko
太阳蓄血证	zh
Greater yang blood accumulation pattern (TM)	en

Definitions

Definition	Language Code
태양병(太陽病)에서 사기(邪氣)가 제때에 제거되지 않으면 열(熱)로	ko

Figure 7.
Shows the detailed concept statistics view for the concept “22.04.420 ‘Greater yang blood accumulation pattern (TM)’” of the ICTM data set. Figure 7a shows the change and note distribution on this concept over time. The distribution of changes across users for the displayed concept can be seen in Figure 7b. A list of assigned communities is displayed in Figure 7c (empty). Parents of the currently browsed concept are listed in Figure 7d. As ICTM features a set of different languages, the various titles and definitions for each concept are depicted in Figures 7e and 7f.

NIH-PA Author Manuscript

NIH-PA Author Manuscript

NIH-PA Author Manuscript

Table 1

A list of features, used to weight nodes in the concept network. Per feature, we highlight corresponding questions that could be tackled with this kind of visualization. Additionally, a list of author specific concept network features, used to weight nodes in author specific network visualizations, and their corresponding research questions are provided.

Feature	Question addressed
Concept Network	
Changes and Notes History	
Number of changes and/or notes	Which are the highly edited/discussed areas in the ontology?
Changes and notes	Which are the highly active areas in the ontology?
Distinct authors of changes/notes	Which concepts attract many different authors?
Authors Gini coefficient	Which concepts are edited more “democratically”, i.e., in a more evenly distributed manner? Contrarily, which are the areas/concepts that are dominated by many changes of a single author?
Overrides	Which concepts cause most disputes (i.e. have the highest amount of changes performed on the same properties of a concept)?
Edit sessions	Which are the highly active areas (with aggregated consecutive changes of the same property by the same author being 1 edit session)?
Distinct authors by property	Which concepts have many properties that are edited by many different authors?
Network Features	
Number of parents/children	Which concepts have many parents? (This is particularly interesting in the case of ICD-11, as multiple parents were not possible in ICD-10 and are therefore introduced gradually.) Which concepts have many children? (i.e. Number of parents/children in the ontological structure)
Depth in network	Which concepts are at what levels in the ontological structure (i.e. what is the shortest path of each concept to the root concept)?
Betweenness centrality (directed/ undirected), Pagerank, Closeness centrality	What are central/popular concepts in the ontology, when looking at different attributes of the network structure of the ontology?
Number of changes by community	How many changes of a concept did each community performed?
Number of titles/definitions and language codes	How many different titles, definitions or language codes are available for each concept?

Author Specific Concept Network	
Changes and Notes History	
Number of changes and/or notes	What concepts or areas of the ontology did a user **either** edit **or** comment frequently? What concepts or areas of the ontology did a user **both**, edit **and** annotate frequently?

NIH-PA Author Manuscript

NIH-PA Author Manuscript

NIH-PA Author Manuscript

Table 2

An excerpt of usability issues identified in the Heuristic Evaluation, manually filtered by significance. The issues are ranked according to the average severity assigned by the three evaluators.

Short Title	Description
Interpretation of results	It is very hard to interpret the currently browsed features (see Table 1) of the network visualizations as no information about the actual meaning (or interpretation) of chosen features is provided.
Details of initial concept network visualization unclear	Directly after login - when being confronted with the initial ontology concept network visualization - it is unclear what parts of the ontology are represented as nodes, edges, colors and diameter of the nodes.
Unreadable hover text	When the mouse is hovering over a node in the graphical representation, the hover text is unreadable if the hovered node has many children and parents.
Wrong vocabulary for Audience	Ontology experts might not be familiar with network analysis measures and vocabularies.
Can't move/drag nodes	The network interface should support “drag & drop” for nodes, as there are bigger nodes that sometimes conceal smaller underlying nodes.