Visualizing User Editing Behavior in Collaborative Ontology-Engineering Projects

Simon Walk[1], Tania Tudorache[1] and Mark A. Musen[1]

[1]Stanford Center for Biomedical Informatics Research

October 17, 2016

Motivation

Over the last decade, ontologies have become the mainstay in the biomedical domain.

Motivation

Over the last decade, ontologies have become the mainstay in the biomedical domain.

- New and complementing areas of application

Motivation

Over the last decade, ontologies have become the mainstay in the biomedical domain.

- New and complementing areas of application
- Increased complexity & size

Motivation

Over the last decade, ontologies have become the mainstay in the biomedical domain.

- New and complementing areas of application
- Increased complexity & size

For example, ICD-11 consists of roughly $50,000$ classes.

Motivation

Over the last decade, ontologies have become the mainstay in the biomedical domain.

- New and complementing areas of application
- Increased complexity & size

For example, ICD-11 consists of roughly $50,000$ classes.

- Highly specialized knowledge
- Many different areas of expertise

Motivation

Over the last decade, ontologies have become the mainstay in the biomedical domain.

- New and complementing areas of application
- Increased complexity & size

For example, ICD-11 consists of roughly $50,000$ classes.

- Highly specialized knowledge
- Many different areas of expertise

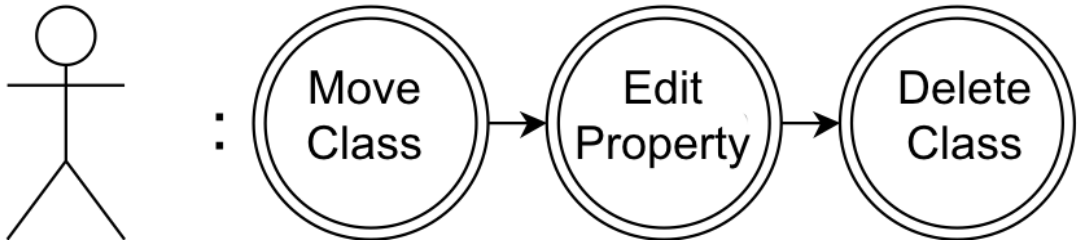
While online collaborative projects have become common, **the processes that drive these collaborations are still not well understood.**

Dataset Characteristics

Table: Characteristics of the ICD-11 and ICTM datasets used in our analyses.

	ICD-11	**ICTM**
Classes #	48,771	1,506
Changes #	439,229	67,522
Users #	109	27
First change date	2009/11/18	2011/02/02
Last change date	2013/08/29	2013/07/17
Editing period (ca.)	4 years	2.5 years

(Sequential) Change-Type Sequences



Types of Changes

Table: Listing of all 15 change-type actions in the change-logs.

Change Type	Description
Add Condition	A restriction is added to a class.
Add Direct Type	A direct type is added to an entity.
Add Property Value	A new value is added to a property.
Create Class	A new class is created.
Create Reference	A new reference is created.
Delete Class	A class is deleted.
Delete Condition	A restriction is deleted from a class.
Delete Property Value	A property value is deleted.
Edit Property Value	A property value is edited.
Import Property	A property value is imported from an external ontology.
Move Class(es)	One or more classes are moved in the class hierarchy.
Remove Superclass	A superclass of a class is removed.
Replace Reference	A reference is replaced.
Retire Class	A class is retired.
BREAK	30 minutes of inactivity between two actions.

Modeling Editing Behavior

Using Markov chains

Modeling Editing Behavior

Using Markov chains

- State space S, listing all possible states $s_1, s_2, ...s_n \in S$ with $|S| = n$.

Modeling Editing Behavior

Using Markov chains

- State space S, listing all possible states $s_1, s_2, ...s_n \in S$ with $|S| = n$.
- Matrix W, with each element w_{ij} representing the number of transitions between states s_i and s_j.

Modeling Editing Behavior

Using Markov chains

- State space S, listing all possible states $s_1, s_2, ...s_n \in S$ with $|S| = n$.
- Matrix W, with each element w_{ij} representing the number of transitions between states s_i and s_j.
- Transition matrix P with p_{ij} listing the probability to go from state s_i to s_j and $\sum_j p_{ij} = 1$.

Modeling Editing Behavior

Using Markov chains

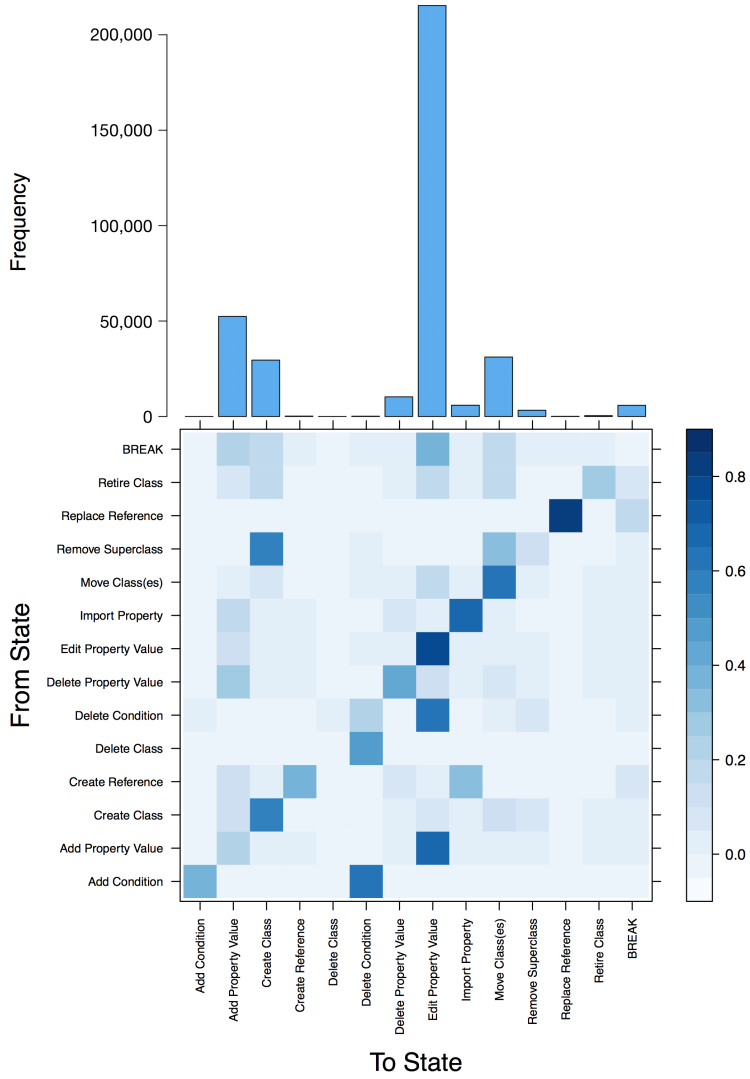
- State space S, listing all possible states $s_1, s_2, ...s_n \in S$ with $|S| = n$.
- Matrix W, with each element w_{ij} representing the number of transitions between states s_i and s_j.
- Transition matrix P with p_{ij} listing the probability to go from state s_i to s_j and $\sum_j p_{ij} = 1$.

First-order Markov chain (Markovian property):

$$P(X_{t+1} = s_j | \underbrace{X_1 = s_{i_1}, ..., X_{t-1} = s_{i_{t-1}}, X_t = s_{i_t}}_{\text{all previous transitions}}) =$$

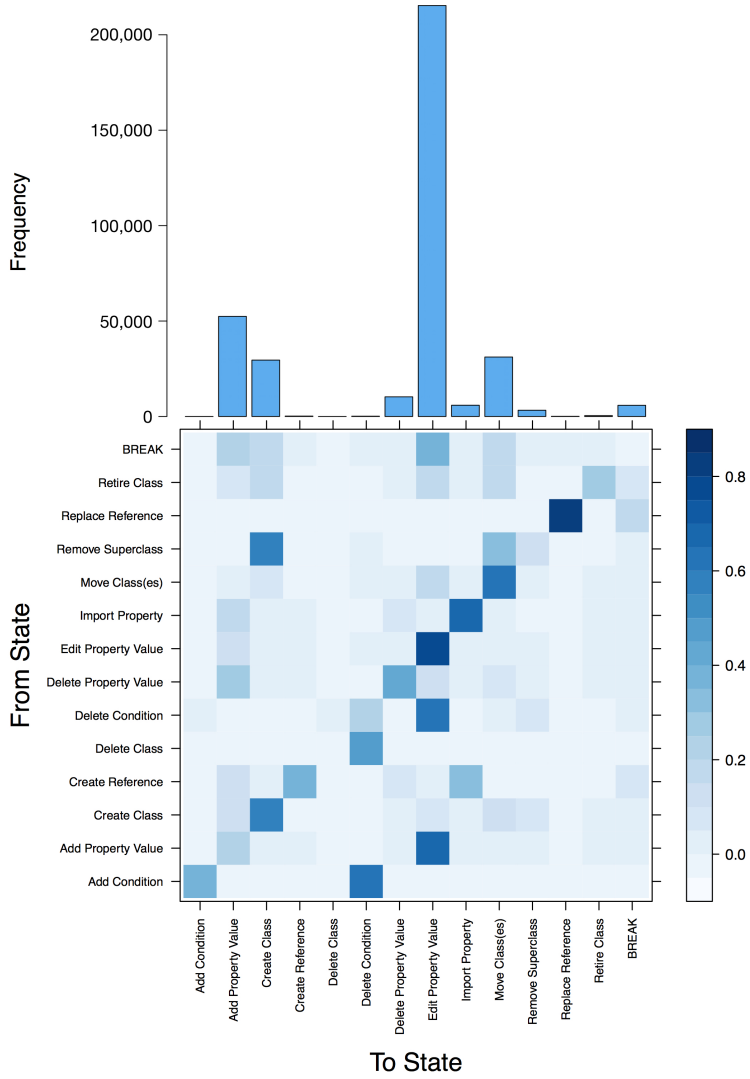
$$P(X_{t+1} = s_j | \underbrace{X_t = s_{i_t}}_{\text{current transition}}) = p_{ij}$$

Fitted Markov Models

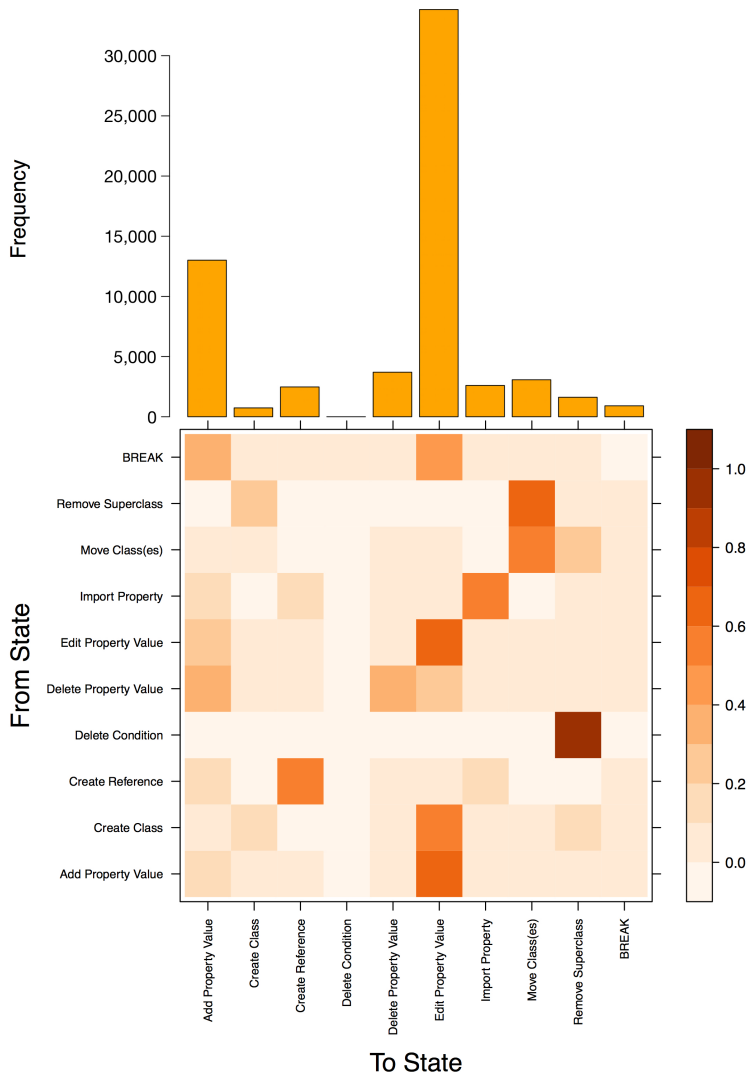


(a) ICD-11

Fitted Markov Models



(c) ICD-11



(d) ICTM

Comparing Interaction Behaviors

To compare **absolute differences** between the two projects, we calculate Q_{abs} as:

$$W_{abs} = W_{ICD-11} - W_{ICTM} \tag{1}$$

and normalize each row of W_{abs} with its $\ell 1$-norm to get Q_{abs}.

Comparing Interaction Behaviors

To compare **absolute differences** between the two projects, we calculate Q_{abs} as:

$$W_{abs} = W_{ICD-11} - W_{ICTM} \tag{1}$$

and normalize each row of W_{abs} with its $\ell 1$-norm to get Q_{abs}.

To compare the **relative importance** of transitions, we calculate Q_{rel} as:

$$Q_{rel} = P_{ICD-11} - P_{ICTM}. \tag{2}$$

Comparing Interaction Behaviors

To compare **absolute differences** between the two projects, we calculate Q_{abs} as:

$$W_{abs} = W_{ICD-11} - W_{ICTM} \tag{1}$$

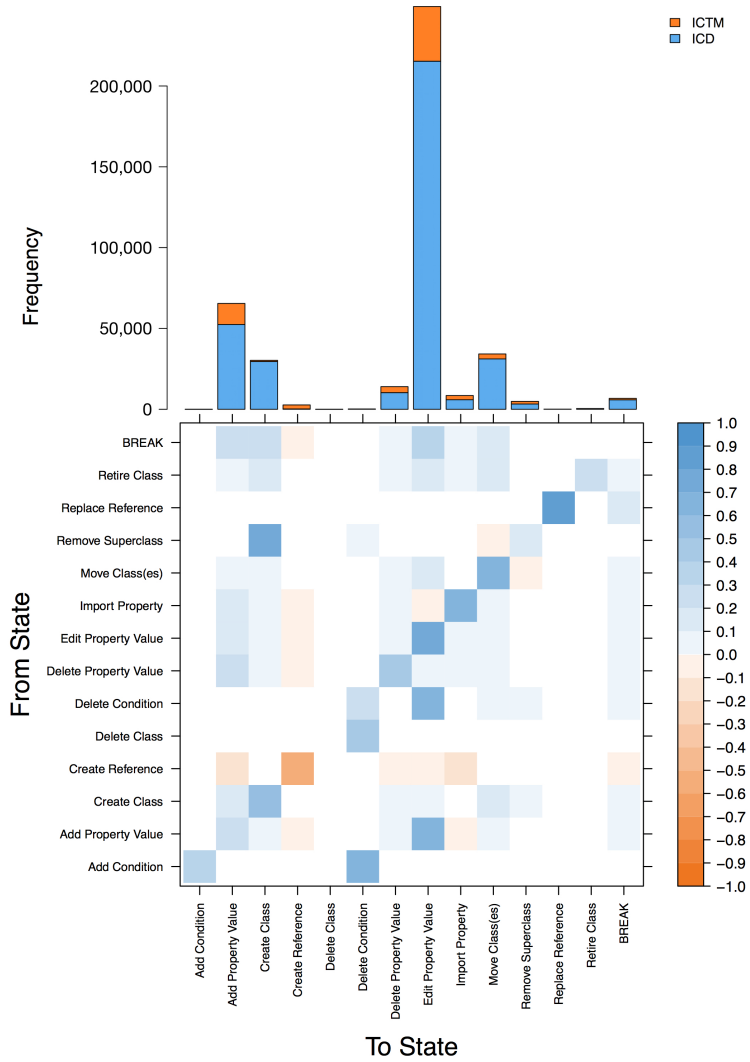
and normalize each row of W_{abs} with its $\ell 1$-norm to get Q_{abs}.

To compare the **relative importance** of transitions, we calculate Q_{rel} as:

$$Q_{rel} = P_{ICD-11} - P_{ICTM}. \tag{2}$$

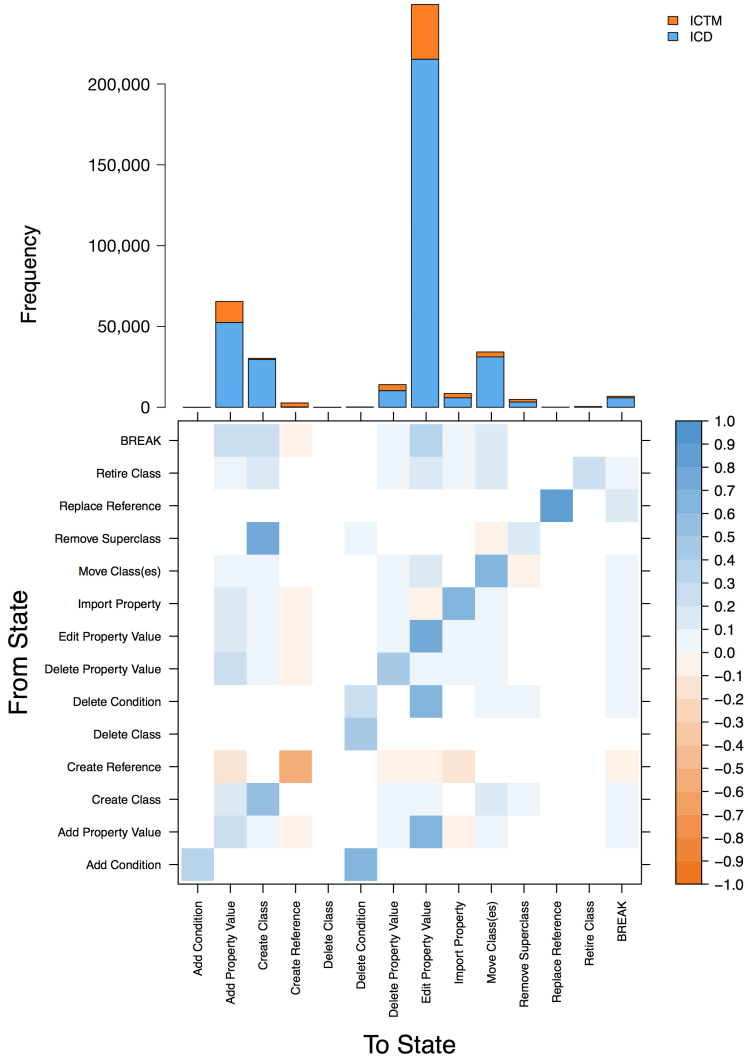
- Transition probabilities range from -1 to 1.

Visualization of Differences

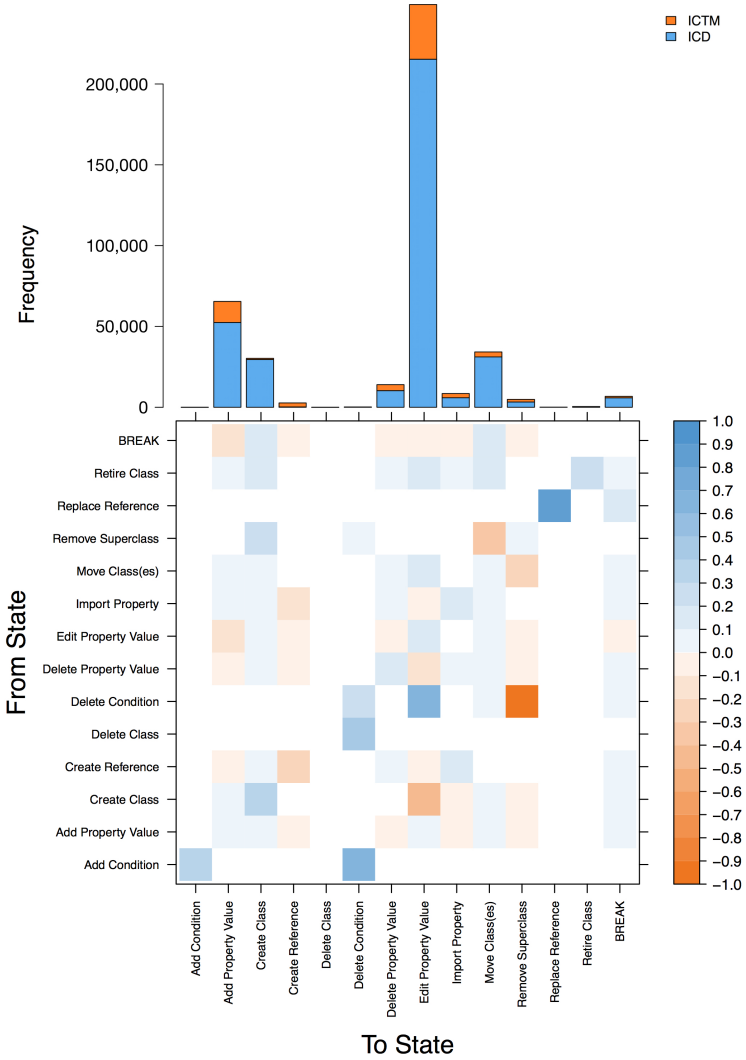


(a) Absolute differences

Visualization of Differences



(a) Absolute differences



(b) Relative differences

Conclusions & Future Work

Conclusions

- We have presented new insights into the editing behavior and workflows of users, not directly visible from static data.

Conclusions & Future Work

Conclusions

- We have presented new insights into the editing behavior and workflows of users, not directly visible from static data.
- The visualizations can be used to identify and analyze differences in the workflows between projects (e.g., different tools or guidelines).

Conclusions & Future Work

Conclusions

- We have presented new insights into the editing behavior and workflows of users, not directly visible from static data.
- The visualizations can be used to identify and analyze differences in the workflows between projects (e.g., different tools or guidelines).

Future Work

- Making the visualizations interactive (e.g., dynamically aggregate similar types of changes into abstract classes of changes).

Conclusions & Future Work

Conclusions

- We have presented new insights into the editing behavior and workflows of users, not directly visible from static data.
- The visualizations can be used to identify and analyze differences in the workflows between projects (e.g., different tools or guidelines).

Future Work

- Making the visualizations interactive (e.g., dynamically aggregate similar types of changes into abstract classes of changes).
- Use dynamic grouping to visualize higher-order Markov chains, and avoid visual clutter due to the increased number of states.

Conclusions & Future Work

Conclusions

- We have presented new insights into the editing behavior and workflows of users, not directly visible from static data.
- The visualizations can be used to identify and analyze differences in the workflows between projects (e.g., different tools or guidelines).

Future Work

- Making the visualizations interactive (e.g., dynamically aggregate similar types of changes into abstract classes of changes).
- Use dynamic grouping to visualize higher-order Markov chains, and avoid visual clutter due to the increased number of states.
- Compare the editing behavior of users across different ontology-development tools to assert the influence of the tool on the editing behavior.

Questions?

Thanks!

Thanks!

NSF Award#: IIS-1622628

IIS: Travel Fellowships for Students from U.S. Universities to Attend ISWC 2016

Timeline: March 1, 2016 – February 28, 2017