Investigating Collaboration Dynamics in Different Ontology Development Environments

Marco Rospocher

DKM, Fondazione Bruno Kessler
rospocher@fbk.eu

Tania Tudorache and Mark Musen

BMIR, Stanford University
tudorache@stanford.edu, musen@stanford.edu

Introduction

- Understanding processes and dynamics behind the collaborative development of ontologies is important

  - for Ontology tool engineers

    - to understand how to optimize their tools to make the work of the users more straightforward and effective

  - for Ontology project managers

    - to obtain tools and metrics to assess and monitor the development status and the quality of the ontology under their responsibility

FONDAZIONE
BRUNO KESSLER

STANFORD
UNIVERSITY

Our Contribution

- We conducted some exploratory investigations on
    - the way people edit an ontology in collaborative settings
    - the role of discussion activities in collaborative ontology development
- Novelty:
    - two different ontology development frameworks
    - discussion activities
- Key Aspects:
    - analysis based on truly objective data
    - five real ontology development projects

Investigations

I1. Is the editing process localized?

I2. Is the formalization of an ontology entity truly collaborative?

I3. Are discussed ontology entities actually discussed by two or more users?

I4. Are highly discussed ontology entities also highly edited?

I5. Do users tend to edit more than to discuss?

FONDAZIONE
BRUNO KESSLER

STANFORD
UNIVERSITY

Tools: protégé

- A collaborative ontology authoring tool for the Web

- Form-based mechanism

- Extensive collaboration support

  - tracking of all changes that users perform in a structured log

  - notes and (threaded) discussions

  - Ontology entity/branch watch mechanism (with email notifications)

  - highly configurable access policies

- In this study we used iCAT

  - a custom configuration of WebProtégé used by medical experts

Tools: protégé

Tools: MoKi
the Modelling WiKi ---

- A collaborative wiki-based tool for modeling ontological and procedural knowledge

- Form-based editing

- Different ways to navigate the ontology under development
  - class hierarchy tree / list in a tabular form all entities defined in the ontology / search for a specific entity

- Support for user collaboration
  - discussions, by means of talk pages
  - watchlists and notifications
  - recent activity awareness features

Tools: MoKi
the Modelling WiKi ---

| page | discussion | unstructured | lightly-structured | fully-structured | edit | history | d |

MoKi
the Modelling WiKi ---

Edit
- Add/Edit a Concept
- Add/Edit an Individual
- Add/Edit a Property
- Add/Edit a DataType Property

List
- List all Concepts
- List all Individuals
- List all Properties
- List all DataType Properties

Visualize
- IsA Browser
- Individuals Browser

Collaboration
- Recent Changes
- New Discussions
- Recent Discussions
- Users' Contributions
- My watchlist

Lightly-structured: Mountain

is a

Every Mountain is a `Landform` [Remove]

[Add another isa axiom]

has part

[Add another has part axiom]

Properties

Subject	Property	Object	
Mountain	hasLocation	GeograficalPlace	[Remove]

[Add another property]

Verbalized

- Every Mountain is something that is not a Hill and that is not a Plain.
- Everything that is MadeOf by a Mountain is something that is an Earth or that is a Rock.

[Save]

FONDAZIONE BRUNO KESSLER

STANFORD UNIVERSITY

protégé and **MoKi** the Modelling WiKi ---

- Main similarities
  - web-based
  - functionalities for supporting both editing and discussion
  - user editing mainly via forms
  - similar collaboration features (e.g., notification, watchlist, history)

- Main differences
  - granularity and the modality of editing and discussion activities
  - navigation and hierarchy awareness
  - discussion awareness

- Motivations for using WebProtègè and MoKi in our study
  - provide detailed change and discussion logs
  - used in several real-world projects

Ontologies

- WebProtégé Ontology Development projects considered:

  - The 11th Revision of the International Classification of Diseases (ICD)

    - led by the World Health Organization (WHO)

    - a taxonomy and descriptions of diseases used in United Nations countries

  - The International Classification of Traditional Medicine (ICTM)

    - led by the World Health Organization (WHO)

    - standard terminology and classification for diagnoses and interventions in Traditional Medicine

    - 4 languages covered: English, Chinese, Japanese and Korean

Ontologies

- MoKi Ontology Development projects considered:

    - Organic Agriculture (OA)

        - classify educational material in a multilingual web-portal containing organic agriculture and agro-ecology resources

        - 15 languages

    - Viticulture (Vit)

        - concepts related to the science, production, and study of grapes

        - no discussion functionalities

    - Motivation and Emotion (ME)

        - motivational and emotional aspects of the learning process in pedagogy

        - educational material and the interventions to be used for facing motivational or emotional difficulties

Ontologies

|	ICD	ICTM	OA	Vit	ME
Developed with	*protégé*	*protégé*	MoKi the Modelling WiKi	MoKi the Modelling WiKi	MoKi the Modelling WiKi
No. of ontology: • classes • individuals • properties	• 50,609 • 228,629 • 228	• 1,511 • 18,364 • 219	• 284 • 81 • 31	• 481 • 0 • 0	• 72 • 13 • 13
No. of active users	109	23	10	3	3
No. of edits	331,147	40,840	2,915	2,227	407
No. of discussions	71,371	1,726	452	0	52
Status	ongoing	ongoing	ongoing	ongoing	completed
Development period (ca)	42 months	30 months	5 months	3 months	1 week
Used for investigations	All	All	All	I1, I2	All

Results

- We tested whether users, after editing a class A, tend to edit another class B closely or semantically related to the previous one
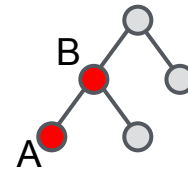
- Six cases considered:
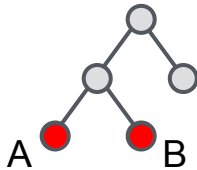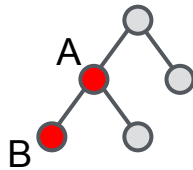
siblings   child   parent
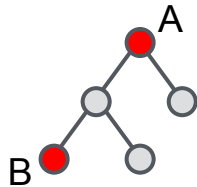
descendant   ancestor   none
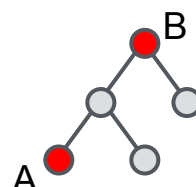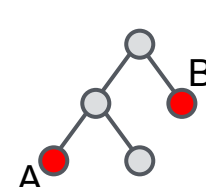
- We counted the number of these occurrences, normalizing over the total number of cases

FONDAZIONE BRUNO KESSLER

STANFORD UNIVERSITY

Results

- Most of the times (60% to 73%), next edited entity is a sibling, a child or a parent

    - Exception: OA

        - strong multilingual focus

        - editing following the alphabetically-sorted list of concepts
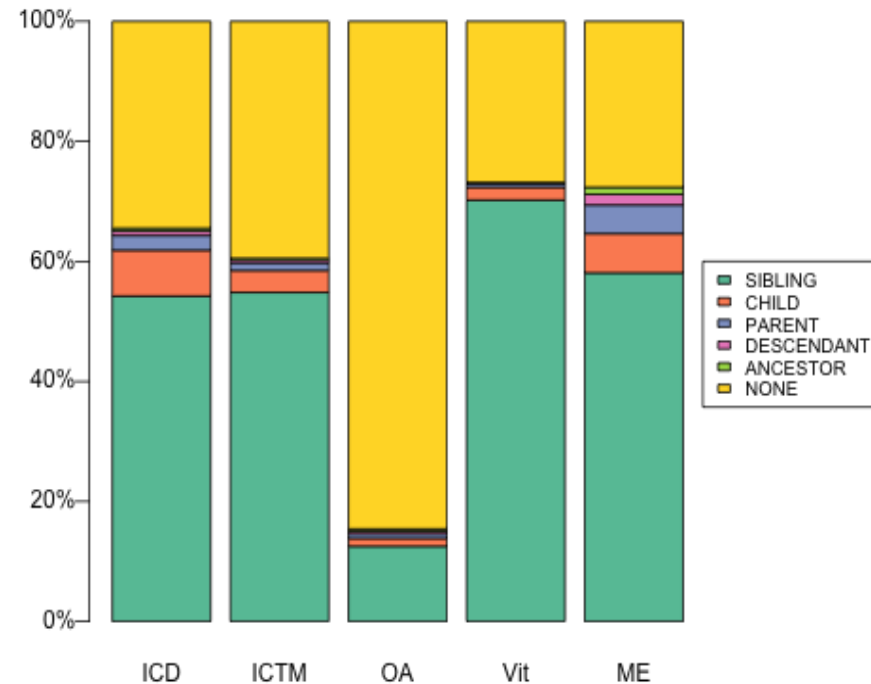
- Outcome: users tend to work locally on the ontology

    - may be due to class navigation functionalities, similar yet different in the tools

        - (to be further investigated) Do tool functionalities impact the way people perform their editing activities?

Results

- We examined how many distinct users usually edit an ontology entity, whether a class, individual, or property

- We classified ontology entities in three categories

  - edited by only one user

  - edited by two distinct users

  - edited by three or more distinct users

FONDAZIONE
BRUNO KESSLER

STANFORD
UNIVERSITY

Results

- Most of the ontology entities (75% to 96%) edited by at most 2 users

- Exception: OA
  - 65% of the entities edited by at least five distinct users
  - strong multilingual focus
  - rather low entities / user ratio (~40), multiple users editing activities more likely to occur
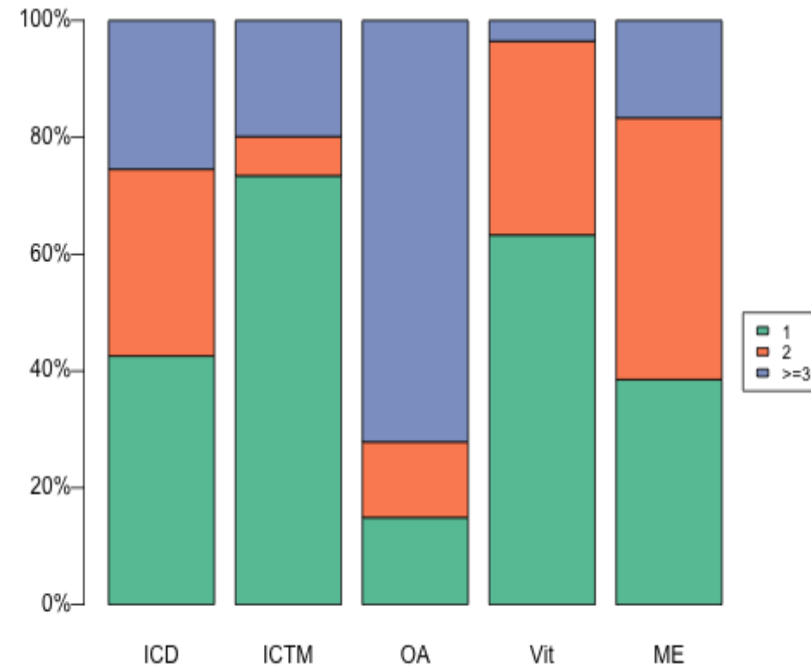
- This kind of analysis may provide useful insights also to ontology project managers:
  - to detect entities having a very few number of editors (may require some intervention)

Results

- We examined how many distinct users usually discuss an ontology entity, whether a class, individual, or property

- Similarly to I2, we classified ontology entities in three categories

  - discussed by only one user

  - discussed by two distinct users

  - discussed by three or more distinct users

FONDAZIONE BRUNO KESSLER

STANFORD UNIVERSITY

Results

I3: Are discussions truly collaborative?

- ICD and ICTM: most of the ontology entities (~ 91%) are discussed by a single users

- OA and ME: most of the ontology entities (resp., 97% and 75%) are discussed by at least 2 users
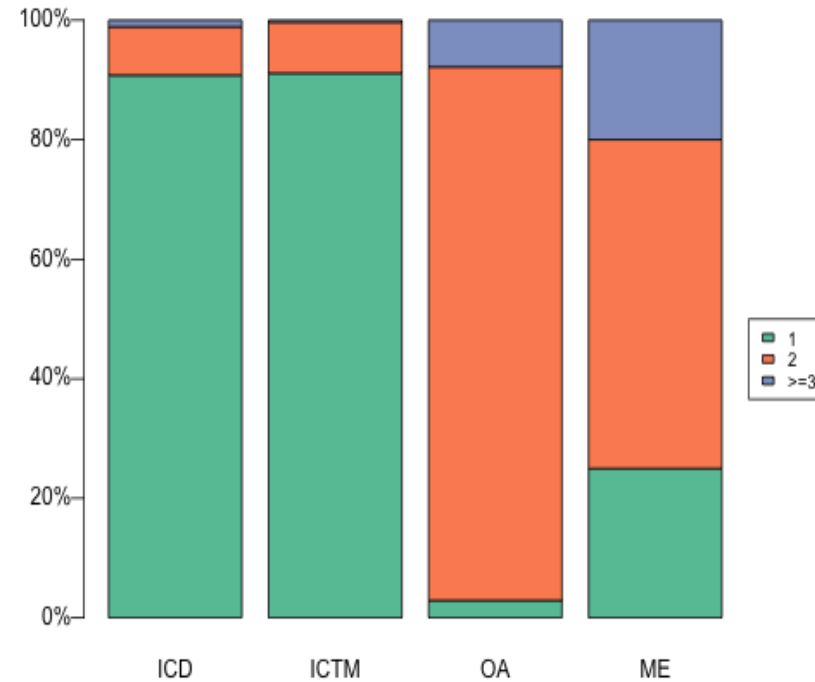
- Possible explanations:
  - difference in size of the ontology and the number of users
    - entity / user ratio (ICD: ~382 and ICTM: ~171; OA: ~40 and ME: ~33)
  - different discussion-awareness support provided by the tools
  - ICD and ICTM have used the notes mechanism not only for discussion, but mostly for notes or additional documentation

Results

- We examined a possible correlation between the discussion activities and the editing activities on an entity

- We classified the ontology entities in two groups:
  - those having at least two distinct users discussing each of them
  - those having zero or at most one user discussing them

- We then computed
  - the average/median of the number of editing activities on an entity

Results

I4: Are highly discussed entities also highly edited?

Number of distinct users

average

median

■ pages discussed by at most one user

■ pages discussed by two or more users
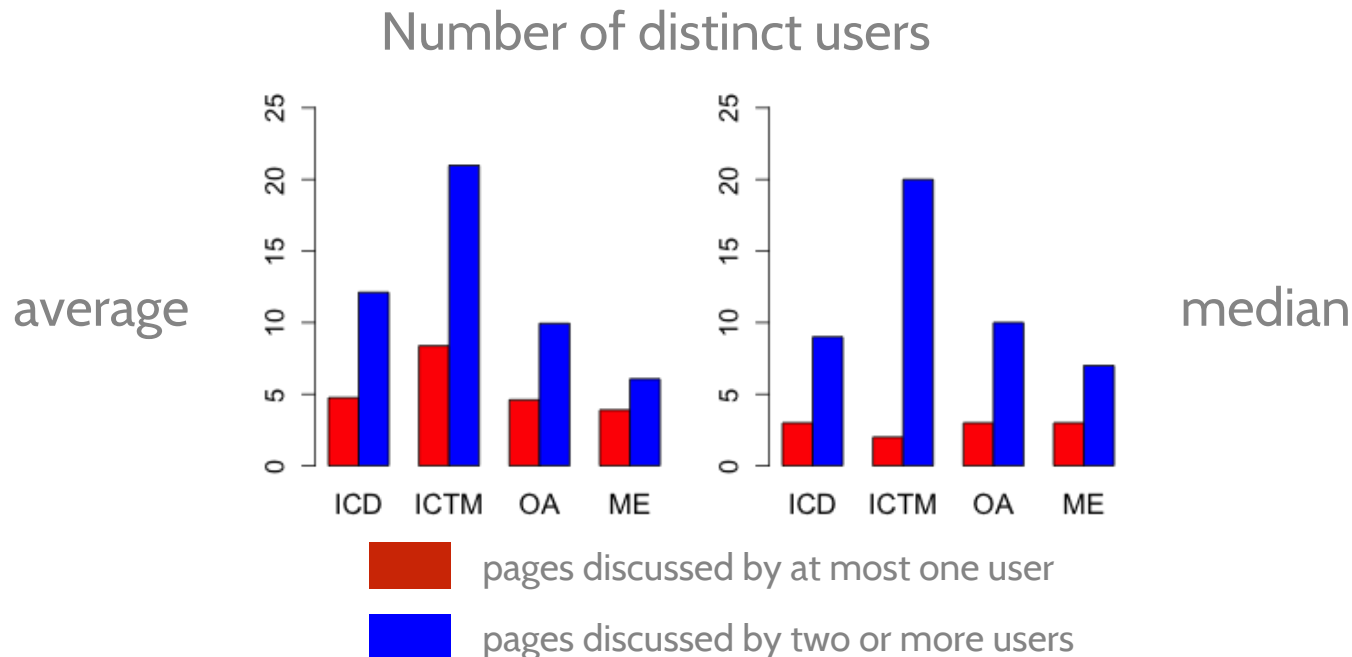
- More editing activities on the set of entities having at least two users discussing them (p<0.05 with Wilcoxon test)

- Encouraging and facilitating the use of discussion support functionalities, to favor the increase of editing activities?

  - to be further investigated

FONDAZIONE BRUNO KESSLER

STANFORD UNIVERSITY

Results

- We examined whether users tend to perform more editing activities than discussion activities
- For each user, we counted the number of editing activities and discussion activities performed

FONDAZIONE
BRUNO KESSLER

STANFORD
UNIVERSITY

Results

editings ----- discussions
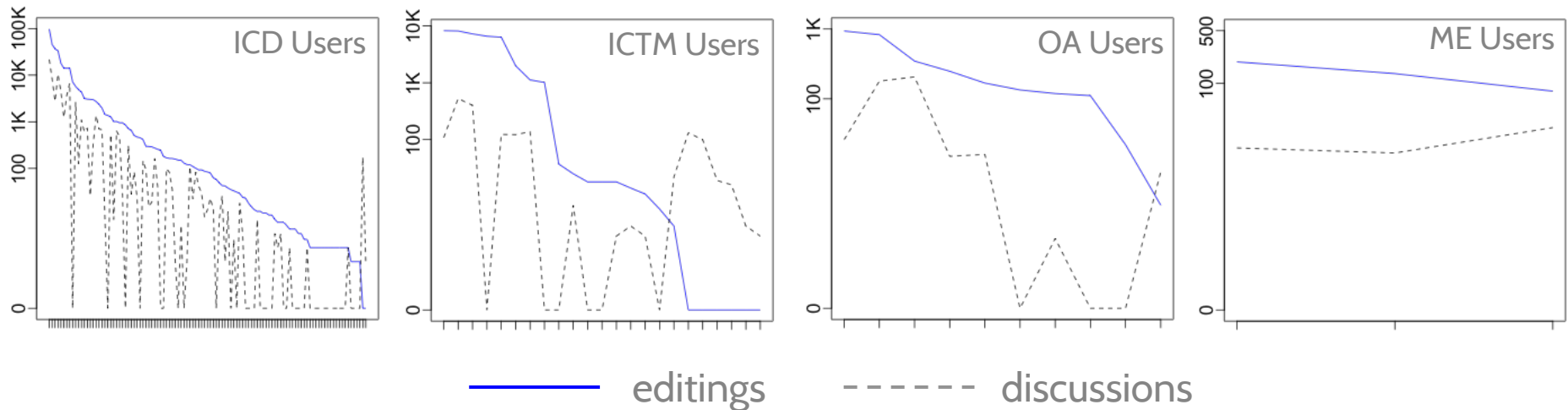
- In most of the cases, users performed much more editing activities than discussion activities
  - a few exceptions in which the contrary holds → they may indicate the existence of different types of users
    - users who prefer sharing thoughts and opinions
    - users that mainly review and comment work performed by others

Limitations

- We used ontologies developed with WebProtégé and MoKi

  - we plan to perform the analysis with additional tools

- The ontologies we analyzed in our study vary in size and in the number of users participating in their development

  - we will consider additional ontology projects

    - e.g., development of some large ontologies with MoKi, or the modeling of a small focused ontology with WebProtégé

- Our analysis is based exclusively on the logs of the activities performed by users while using the tools

  - we plan to complement our analysis with additional experimental study techniques covering activities taking place outside the modelling tool

Conclusions and Future Work

- We investigated the collaborative process behind the development of some ontologies modeled with WebProtégé and MoKi

  - focus: the way users edit the ontology and the role of discussion

- Among the findings:

  - users tend to edit ontology entities that closely related to the previously edited one

  - any ontology entity is edited/discussed by few users (≤ 2)

  - the more an ontology entity is discussed, the more likely it is highly edited as well

  - users tend to edit more than to discuss

FONDAZIONE BRUNO KESSLER

STANFORD UNIVERSITY

Conclusions and Future Work

- The results of our analysis raise some suggestions for ontology tool engineers:

  - offering different ontology browsing functionalities

  - better support discussion activities by enhancing discussion awareness

- Future Work

  - extend our study to consider additional ontology development projects

  - include ontologies developed with other tools

    - a detailed tracking of user editing and discussion activities is required for computing the metrics applied in our study

  - in-depth investigation of the influence of the user interface / tool features on the dynamics of the collaboration processes

Thank you! Questions?

Marco Rospocher

DKM, Fondazione Bruno Kessler

rospocher@fbk.eu

Tania Tudorache and Mark Musen

BMIR, Stanford University

tudorache@stanford.edu, musen@stanford.edu

http://moki.fbk.eu

http://webprotege.stanford.edu