# Exploiting Online Discussions in Collaborative Distributed Requirements Engineering

Itzel Morales-Ramirez, Matthieu Vergne, Mirko Morandini, **Anna Perini**, and **Angelo Susi**

Software Engineering Research Unit, Fondazione Bruno Kessler (I)
Email: {perini,susi}@fbk.eu

8th International iStar Workshop 2015

Motivation

- Context of the work: distributed, collaborative software development, such as in OSS projects

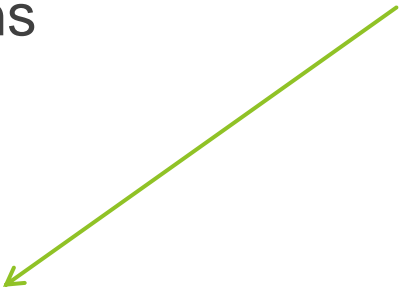- Mailing list discussions are a typical communication channel in OSS
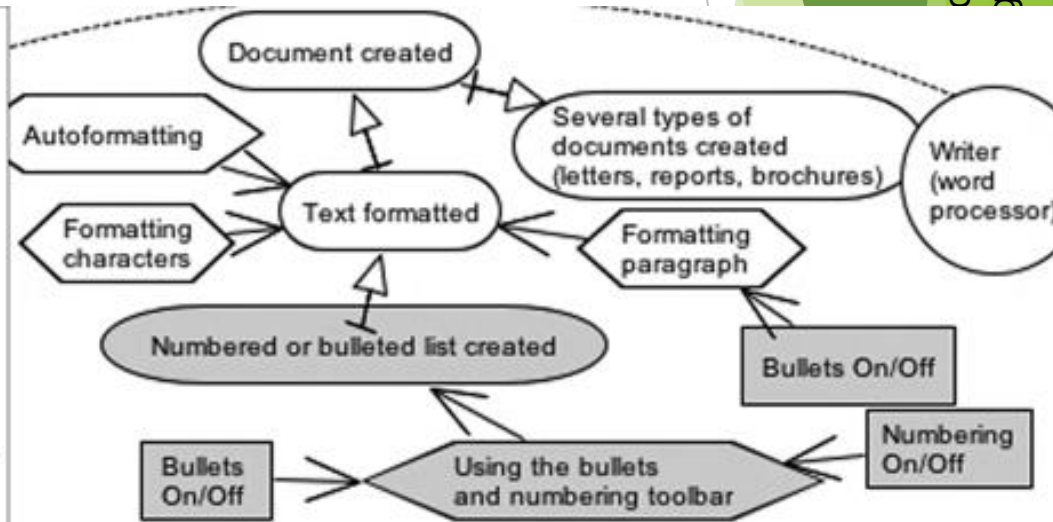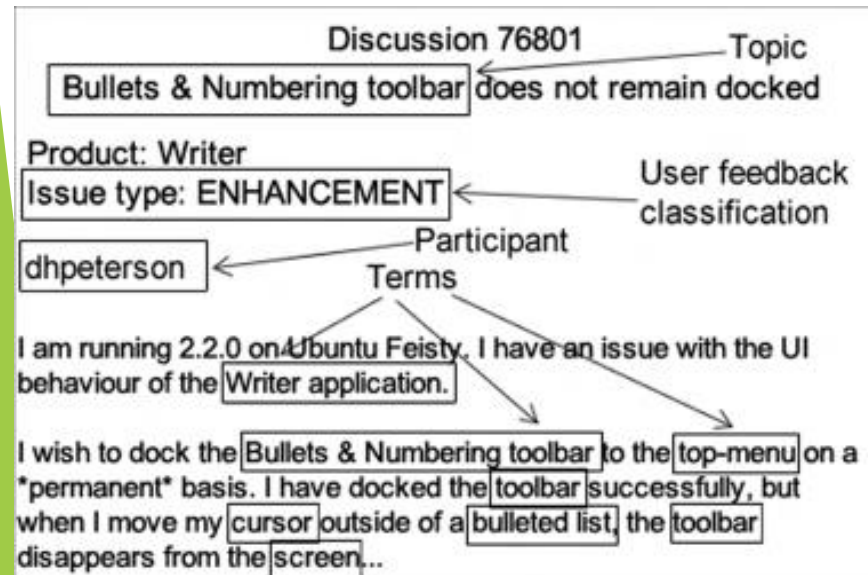
Motivation

- We believe that "expert user feedback" expressed in online discussions can provide relevant **requirements knowledge** for evolving a software application [ref. our work presented at iStar13]

- **Automated support** for extracting relevant information (i.e. indicators of feature requests, or bug) and for identifying and ranking those participants who can contribute key knowledge (i.e. "experts") about a given topic
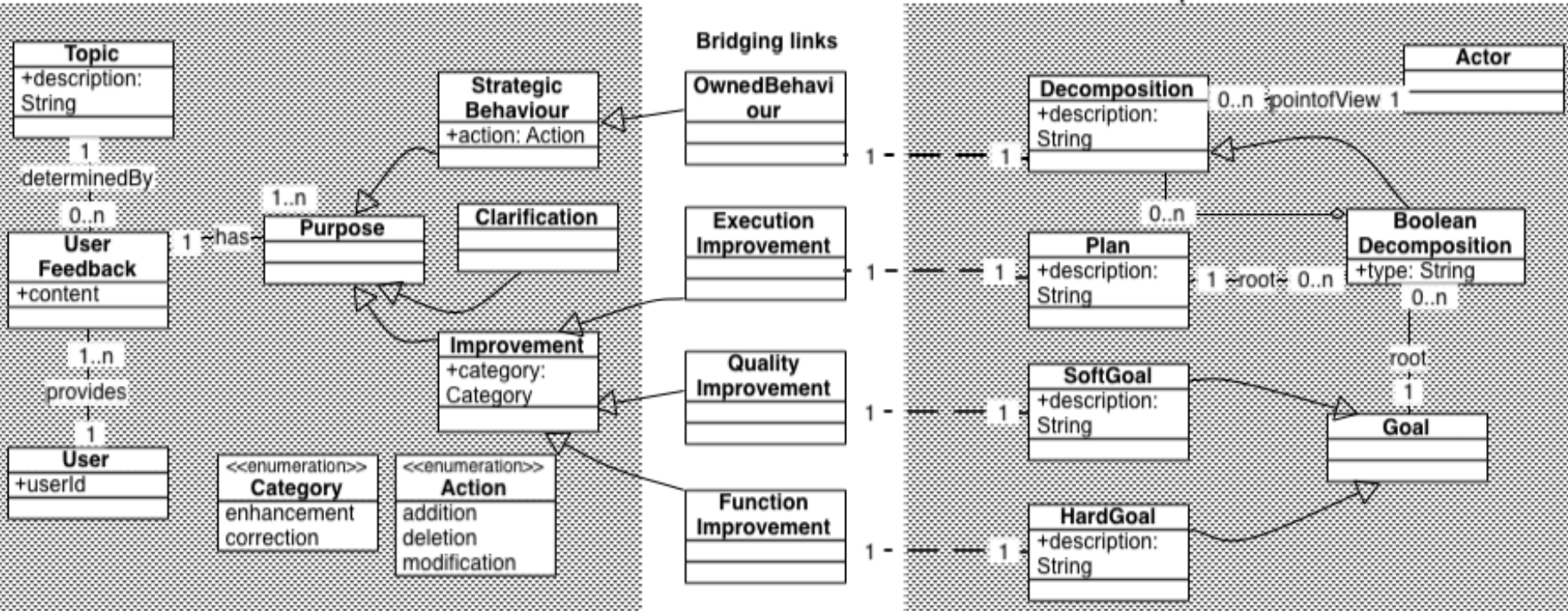
Approach overview

▶ Discussants' knowledge: content in messages

▶ Discussants' intentions: asking, suggesting, stating problems or answering questions

▶ Our approach consider the *content-* and *intent-* dimensions

Terms written in messages such as nouns, adj+nouns

Intention identified by a specific sentence structure

Discovering discussants' intentions

▶ Online discussions can be considered almost synchronous written conversations that can be described in terms of speech-acts, according to the Speech Act Theory [John Searle]

- speech acts are classified according to specific performative verbs, such as *suggest*, *recommend*, *confirm*, and *advise*, etc., which reveal the speakers' **intentions**, such as that of persuading, inspiring or getting a hearer to do

  - E.g. "Open the door, please!"

▶ In our approach, we aim at recognizing those fragments of conversation that contain specific **speech-acts combinations or patterns** (we call this problem requirements knowledge discovery in online discussions)

- Those that are found to be commonly used for expressing feature requests, bugs or clarification requests

Exploiting Online Discussions in Collaborative Distributed Requirements Engineering

Tool-supported annotation of Discussants' Intentions
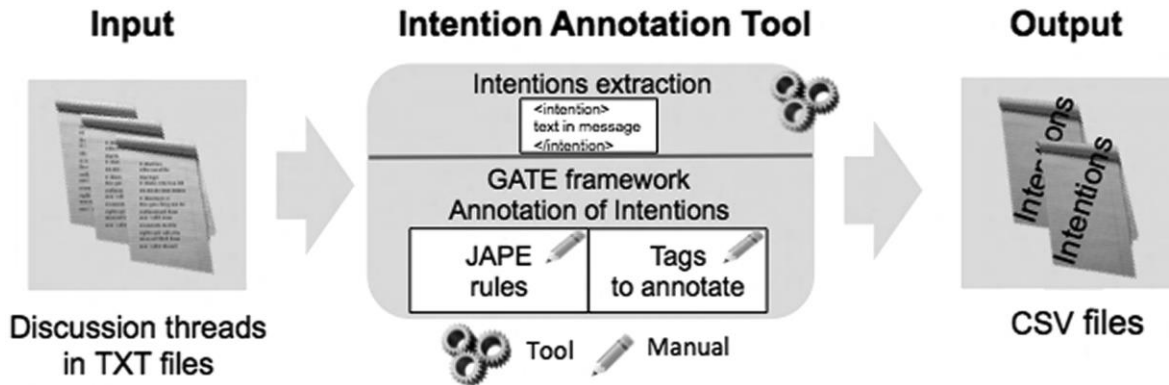
▶ Features to design rules:

 ▶ Length of words

 ▶ Special types of verbs

 ▶ Seed words …

Category	Tag	Rule
Directives	Questions	$<WRB>+<PRP>+<content>+$ "?" $<MD>+<PRP>+<content>+$ "?"
Constantives	Suggestives	$<PRP>+(<MD>)^*+($ "try" $\vert$ "check" $)$ $<PRP>+(<MD>)^*+($ "suggest" $\vert$ "recommend" $)$
	Responsives	$(<PRP>)^* +$ "[Hh]ope" $+<content>+$ "help"

The *Expert Finding* problem

▶ the problem of expert finding in online discussions can be stated as the problem of ranking the *Participants* in those discussions (users and developers participating to the mailing list) according to their expertise on a topic

▶ it can be conceived as a problem of Information Extraction

Approach overview: the process

1. Extract key elements in messages

2. Extract relations between the key elements

3. Combine content- and intent-based information to weight the relations

4. Use Markov Network (MN) to infer the expert stakeholders for a set of topics

Intent

Ranking:

1 s_2

2 s_1
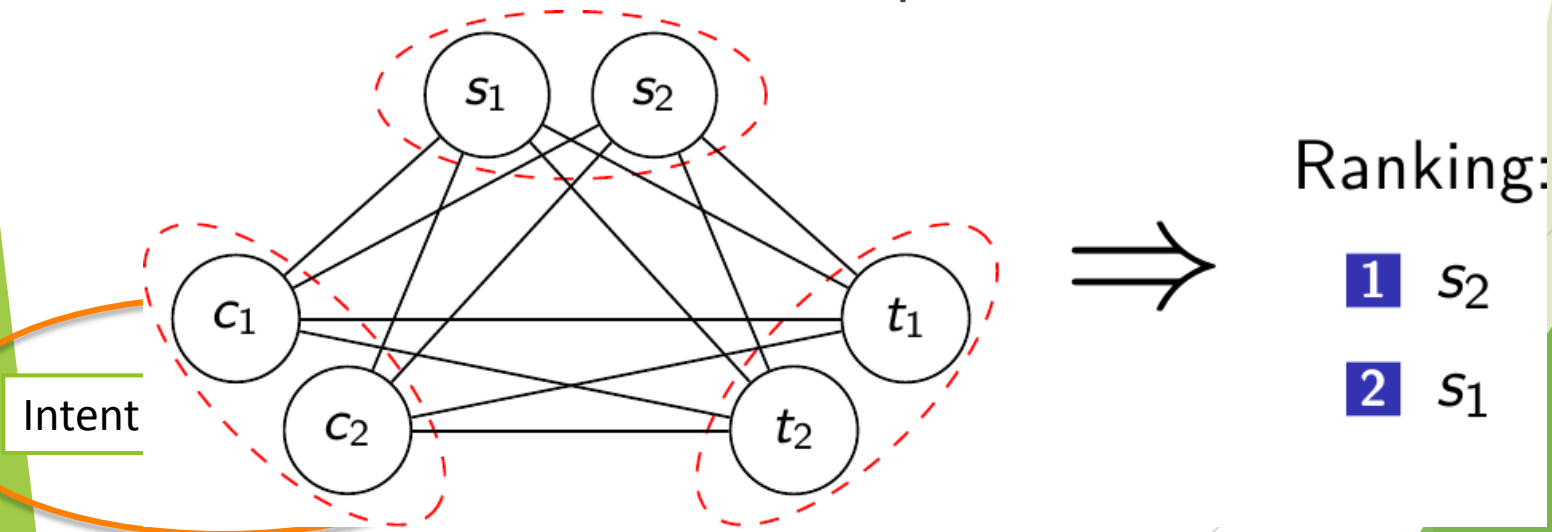
Results so far

▶ Requirements knowledge discovery

  ▶ Extract Speech-Act classification rules using an NLP-based annotation tool (i.e. Using the GATE tool)

  ▶ Perform automated classification of online discussions into Speech-Act / patterns using the machine learning library WEKA

▶ Expert finding problem

  ▶ MN techniques have been applied to OSS discussions of hundreds of sentence, and start revealing stability problems for larger discussions

  ▶ Search-based techniques seem promising for larger discussions

Exploiting Online Discussions in Collaborative Distributed Requirements Engineering

Ongoing and Future work

▶ Ongoing

   ▶ Consolidation of a conceptual framework to performed a speech-act based analysis of online discussions

   ▶ Conceptualisation of expertise's indicators

▶ Future work

   ▶ Compute similarity between phrases in the selected conversation and requirements descriptions (e.g. goal, task labels, or corresponding textual description) to identify the affected requirements

Conclusion

- ▶ online discussions are a source of requirements knowledge (e.g., user's goals and preferences, domain assumptions)

- ▶ NLP, ML, MN and search based techniques have been combined to support the identification of "expert" comments that can help evolve existing requirements model

- ▶ … future work: tool-support for Feedback impact assessment (3)

Exploiting Online Discussions in Collaborative Distributed Requirements Engineering

11
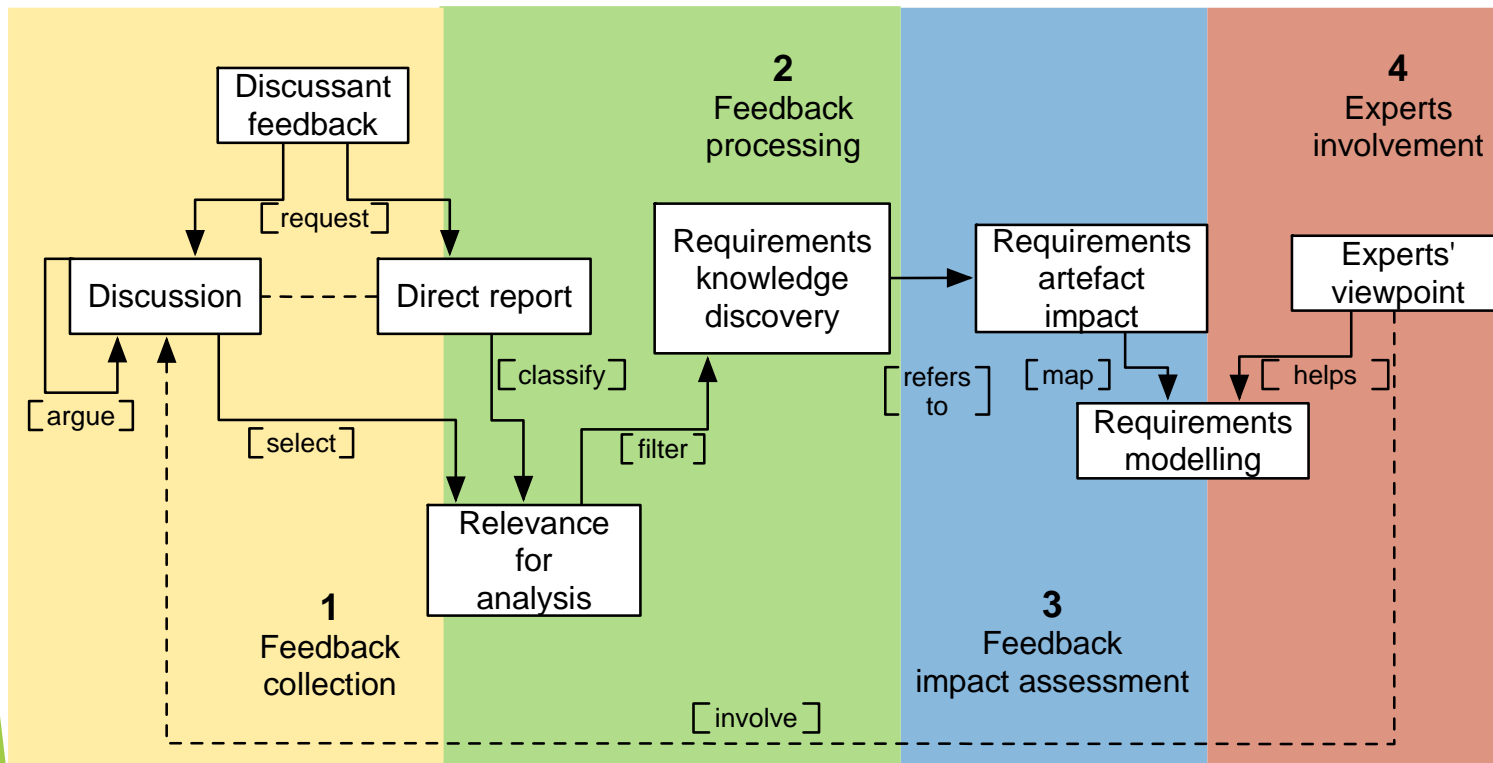
Thank you for your attention!

▶ Questions?