Software Engineering

- Not a stand-alone activity
- People use forums, Stack Overflow, etc. to debug code and find solutions
Social Software Engineering

- Software Engineering has always been social - people work in large teams and collaborate with each other
- Make Software Engineering easier by using the Social Aspects of software development
Codebook (Microsoft Research)

- Social Networking over Code
- Use Source Code artifacts to provide useful information
- Papers
  - Begel, Phang, and Zimmermann, *Codebook: Discovering and Exploiting Relationships in Software Repositories*, ICSE 2010
Codebook

1. Find out who the most relevant engineers (developers, testers, program managers, operations, leads, etc.) are in order to contact them

2. Find an expert to talk to who knows a lot about a feature, API, product or service

3. Get a list of servers, directories and repositories where the related code, bug reports, work items, specifications, etc. are located

4. ...
Codebook
Gail Murphy’s group - University of British Columbia

- Fritz and Murphy, Using information fragments to answer the questions developers ask, ICSE 2010
Developers’ questions

Who is working on what (people specific)
1. Who is working on what?*
2. What are they [coworkers] working on right now?*
3. How much work [have] people done?*
4. Who changed this [code], focused on person?*
5. Who to assign a code review to? / Who has the knowledge to do the code review?
6. What have people been working on?*
7. Which code reviews have been assigned to which person?*
Developers’ questions

Changes to the code (code specific)
8. What is the evolution of the code?
9. Why were they [these changes] introduced?*
10. Who made a particular change and why?
11. What classes has my team been working on?*
12. What are the changes on newly resolved work items related to me?
13. Who is working on the same classes as I am and for which work item?
14. What has changed between two builds [and] who has changed it?*
15. Who has made changes to my classes?
16. Who is using that API [that I am about to change]?
17. Who created the API [that I am about to change]?
18. Who owns this piece of code? / Who modified it the latest?*
19. Who owns this piece of code? / Who modified it most?*
20. Who to talk to if you have to work with packages you haven’t worked with?
21. What classes have been changed?*
22. [Which] API has changed (to see which methods are not supported any more)?*
23. What’s the most popular class? [Which class has been changed most]?*
24. Which other code that I worked on uses this code pattern / utility function?
25. Which code has recently changed that is related to me?
26. How do recently delivered changes affect changes that I am working on?*
27. What code is related to a change?
Developers’ questions

**Broken builds**

32. What caused this build to break? (Which change caused the stack trace?)
33. Who caused this build to break? (Who owns the broken tests?)
34. Who changed the test case most recently that caused the build to fail?
35. Which changes caused the tests to fail and thus the build to break?
Developers’ questions

Other Questions
42. How is the team organized?*
43. Who has made changes to [a] defect?*
44. Who has made comments in defect?*
45. [What is] the collaboration tree around a feature?*
46. Which conversations in work items have I been mentioned?*
Answering such questions

Figure 1: Approach to Answer the Question “What have people been working on?”
• More information - Peppo Valetto’s Guest Lecture on Thursday 12/2
Puzzle

• 29 pairs of spouses
• At least 1 man cheated
• Each woman must determine if her husband was faithful
• 2 nights, 1 question per night per woman
• 2 possible gun shots
• What strategy can the women come up with?
Social Software Engineering

- Using Software Engineering for social solutions to problems
...with a little help from my friends
Wisdom of the Crowds

- Wikis
  - Wikipedia, Wikitravel, Wiktionary, ...
- Social bookmarking
  - Reddit, Digg, Delicious, ...
- Social Games
  - MMOGs, Farmville, ...
- ...

Tuesday, November 30, 2010
Recommendations, recommendations, recommendations!!!

- Things to buy (Amazon)
- Movies to watch (IMDb, Netflix)
- Friends you may know (Facebook)
- Music you may like (Last.FM)
- Jokes you may like (Jester)
- ...

Tuesday, November 30, 2010
Collaborative Filtering

- Filter information based on interests
Collaborative Filtering

- Why filter?
  - To make it more relevant, useful

- What information?
  - Likes, dislikes, preferences

- What interests?
  - Movies, Music, Friends, ...
Collaborative Filtering

- How?
  - Memory-Based
    - Find similarity between users or items
    - Cosine Similarity, Pearson Correlation, Nearest Neighbor, ...
  - Model-Based
    - Use data mining or machine learning algorithms
    - Bayesian networks, Clustering, Latent Semantic Models, ...
Netflix & Netflix Prize

- Movie and TV show rentals
- 10 million subscribers
- 100,000 titles on DVD in 2009
- Announced the Netflix Prize in 2006
Netflix Prize

• Improve the recommendations used by Netflix by 10%

• Grand Prize of US$1,000,000

• Released movie rating data

  • ~100,000,000 movie ratings
  • ~500,000 users
  • ~20,000 movies
Netflix Prize

- Techniques used from varying fields: computer science, statistics, machine learning, psychology
- Won by “BellKor's Pragmatic Chaos” in June 2009
- Team Members - 2 from AT&T Labs, 1 from Yahoo!, 1 from Commendo Research and Consulting GmbH, 2 from Pragmatic Theory
genSpace

• Provide social features in the domain of computational biology and bioinformatics

• Use Collaborative Filtering to provide recommendations to users

• Papers
  • Sheth, Arora, Murphy, and Kaiser, weHelp: A Reference Architecture for Social Recommender Systems, SSE 2010
  • Murphy, Sheth, Kaiser, and Wilcox, genSpace: Exploring Social Networking Metaphors for Knowledge Sharing and Scientific Collaborative Work, SoSEA 2008
genSpace

ALL TOOLS:

Top 3 Most Popular Tools
1: Blast SkyBase
2: SOM Analysis
3: Hierarchical Clustering

Top 3 Most Popular Tools at Start of Workflows
1: Protein Function Annotation Server: MarkUs
2: Anova Analysis
3: ARACNE

Top 3 Most Popular Workflows
1: Hierarchical Clustering,ARACNE,ARACNE,
2: Anova Analysis,Anova Analysis,Anova Analysis,Anova Analysis,Anova Analysis,
3: SOM Analysis,SOM Analysis,SOM Analysis,SOM Analysis,

AN INDIVIDUAL TOOL:

ARACNE Analysis

Total usage rate: 287
Total usage rate at start of workflow: 38
The most popular tool used next to this tool: ARACNE Analysis
The most popular tool used before this tool: ARACNE Analysis

*The ranking of tools and workflows in the "ALL TOOLS" section is based on an exponential time-decay function. The "INDIVIDUAL TOOL" section shows the raw data for a tool usage, which is not exponentially weighted.
genSpace
Social Software Engineering

Swapneel Sheth
swapneel@cs.columbia.edu

COMS W4156: Advanced Software Engineering
Fall 2010