

User Feedback for Continuous Software Evolution

Group 2

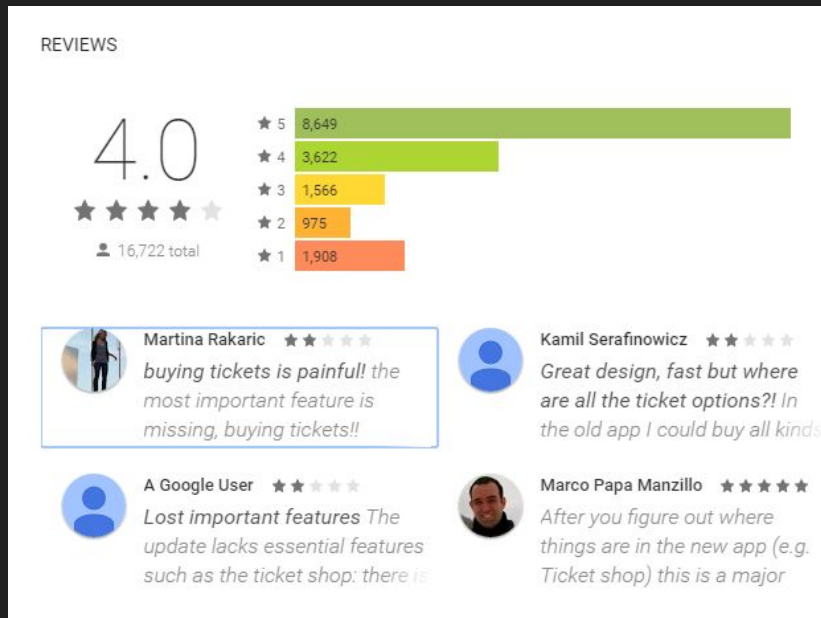
Sebastian Elke & Raphael Rosenast

Agenda

- Introduction
- User Feedback in the App Store
- What Do Twitter Users Say about Software?
- Social Computing Approach in the Course
- Questions?

User Feedback in the App Store

- 6 million apps
- Billions downloads per month
- Feedback
 - Rating
 - Reviews
 - Impact on Downloads



Collected Feedback

1 Million reviews from 1100 applications

- + Helpful Advice
- + Bug Reports
- + Innovative Ideas
- + Feature Request
- + How To
- + ...



- Destructive
- Insults
- Nonsense
- ASCII Art
- Spam
- ...

Study Questions

1. Usage

- When? How often?
- Length? Ratings? Helpfulness?

2. Content

- Types? Patterns?

3. Impact

- Market? Community?

Study Results 1: Usage

- Small group of users
- Shortly after release, quickly decreasing
- More feedback for free apps than for paid apps
- Short messages (Tweet/Mail Length)
- Longer Feedback for paid apps
- Average rating: 4.13 stars
- Only 5% of the Feedback is rated

Study Results 2: Content

#	topic	description	freq.
t1	praise	expresses appreciation	75.36%
t2	helpfulness	scenario the app has proven helpful for	22.45%
t3	feature information	concrete feature or user interface	14.45%
t4	shortcoming	concrete aspect, user is not happy with	13.27%
t5	bug report	bug report or crash report	10.00%
t6	feature request	asks for missing feature	6.91%
t7	other app	reference to other app, e.g. for comparison	3.91%

Study Results 3: Impact

#	topic	Ø rating
t8	recommendation	4.88
t2	helpfulness	4.85
t3	feature info.	4.81
t17	howto	4.80
t1	praise	4.78

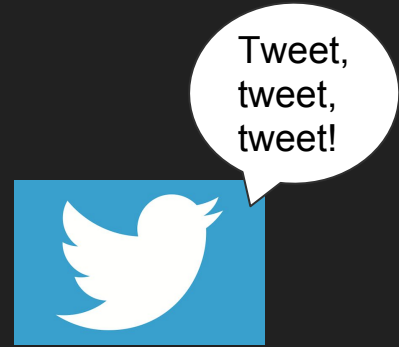
- Positive messages -> better ratings
- Negative feedback -> no user experience info
- “Helpfulness”-Feedbacks are significantly longer
- Feedback rating the App good is more helpful to others

Further Conclusions

- Communication channel among users and developers
 - Lack bidirectional communication features?
 - Improvement: link changes to user feedback
- RE can profit and gain insights
 - How the applications are actually being used
 - Assess the importance of features
 - Collect user experiences by monitoring at runtime
 - Feedback is provided after new releases -> Feature Feedback
- Tools help feedback filtering

Twitter as Source for RE insights

- Highly active community
- Tweets contain relevant information on software
- Direct feedback for continuous evolution of software



Using Tweets for RE (Sample)

- Dataset of 10,986,494 software related tweets
- Randomly selected 1,000 tweets for manual sample checking
- Automated approaches can filter irrelevant tweets with a precision ranging between 0.78 and 0.54
- Machine learning techniques have the capacity to identify and harvest relevant information automatically

Using Tweets for RE

- 25% content related information
- Relevant information is rare
- Almost 70% of tweets contain no relevant information
- Estimate of 9.5% of data was generated by bots

TABLE IV: Manual content analysis results.

Category	Frequency %	Relevance %			Score	Sentiment Interpretation
		Technical	Non-technical	General public		
Feature shortcoming	1.50	100.00	93.33	93.33	-1.07	negative
Feature strength	0.80	100.00	100.00	100.00	0.03	neutral
Feature request	0.10	100.00	100.00	100.00	0.05	neutral
Bug report	0.90	100.00	88.89	88.89	-0.44	neutral
Usage scenario	2.50	84.00	96.00	84.00	0.00	neutral
Hardware constraint	1.10	27.27	54.55	54.55	0.27	neutral
Software constraint	0.0	N/A	N/A	N/A	N/A	N/A
General praise	2.80	96.43	100.00	100.00	1.21	positive
General complaint	1.10	100.00	100.00	100.00	-1.36	negative
Advertisement	28.30	18.37	98.94	98.94	0.15	neutral
Dissuasion	0.40	100.00	100.00	100.00	-0.25	neutral
Question	0.30	66.67	100.00	100.00	0.00	neutral
How to	3.70	94.59	97.30	97.30	0.00	neutral
Feature information	2.50	76.00	100.00	96.00	0.24	neutral
Software price	8.40	7.14	100.00	100.00	-0.04	neutral
Compliance issue	0.0	N/A	N/A	N/A	N/A	N/A
Software extension	0.10	100.00	100.00	100.00	1.00	positive
Other product	5.90	59.32	88.14	88.14	-0.03	neutral
Service	0.00	N/A	N/A	N/A	N/A	N/A
Social interactions	3.60	25.00	55.56	50.00	0.00	neutral
Content related	25.10	8.37	27.49	37.45	0.09	neutral
Job advertisement	0.30	0.00	100.00	100.00	0.33	neutral
Noise	1.00	0.00	0.00	0.00	0.00	neutral
Unclear	9.70	0.00	0.00	0.00	0.04	neutral
Unrelated	15.10	0.00	0.00	0.00	0.01	neutral
Other	8.10	7.41	49.38	44.44	-0.01	neutral

Conclusion for using Tweets for RE

- Direct Feedback
- Community engagement
- Automated analysis required (machine learning)
- Large active community required
- Mainly bad and irrelevant data
- Potential interaction with artificially generated data (bots)

Social Computing Approach in the Course

- Asynchronous Working on the Project
- Collaboration Using Tools (Google Docs/WhatsApp)
- Lack of Controlling Functions

Questions



Questions

- What are the best feedback channels?
- How to prioritize non-structured feedback?