

Online Design Discussion

Rami Akkad

akkad@inf.fu-berlin.de

Betreuer: Christopher Oezbek

01.03.2009

Abstract

This paper summarizes the research of F. Barcellinis work group. It focuses on their published paper "A socio-cognitive analysis of online design discussion in an Open Source Software community" which analyzed and visualized the distributed online discussion space of the Python project community by developing a method to study asynchronous collaborative design activity in there aspects of the online dynamics: Social, thematic temporal and design. Social design studies how roles emerge in discussion, thematic temporal design studies how design related themes evolve in time. The design dynamics is a content-based study of the discussion space. It analyzes how collaborative design work is done by the community. By presenting and discussing the results we get a picture how knowledge is elaborated in this OSS community.¹

¹All pictures are taken from [FB08]. Numbers written in brackets are a reference to the page number of this paper.

Contents

1	Introduction	1
2	Framework / Research Strategy	1
2.1	Social dynamics	3
2.2	Thematic temporal dynamics	3
2.3	Design dynamics	3
3	Methods	4
3.1	Collection of the data	4
3.2	Three analyzed aspects	5
3.2.1	Social view	5
3.2.2	Thematic temporal view	6
3.2.3	Design view	7
4	Results	7
4.1	Social view	7
4.1.1	Global distribution of statuses and involvement in the discussions	7
4.1.2	Quotation activities	9
4.1.3	Position of participants and their used quotation structure	9
4.1.4	Depth of quotation	11
4.2	Thematic temporal view	12
4.3	Design view	14
4.3.1	Design alternatives	14
4.3.2	Design activities in comments	14
4.3.3	Design activities in quotes	15
4.3.4	Sequences of activities	15
5	Discussion	15
5.1	Social dynamics	16
5.2	Thematic temporal dynamics	16
5.3	Design dynamics	17
5.3.1	Activities	17
5.3.2	Sequence of activities	17
6	Summary / Future	17

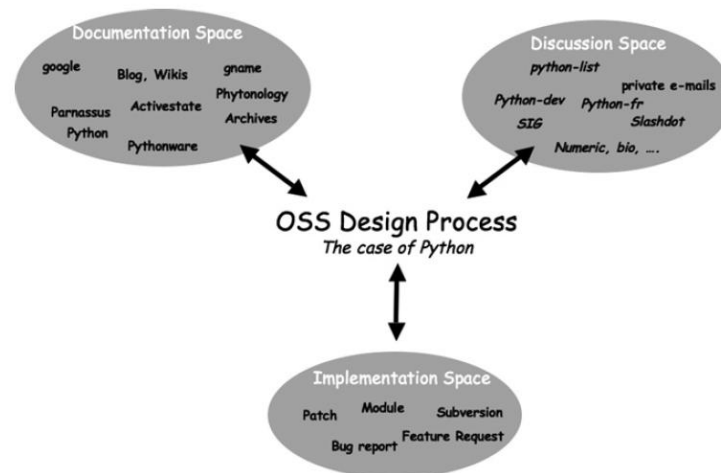
1 Introduction

Open Source Software (OSS) is a constantly growing market in the software industry. OSS projects like Linux, Mozilla or Apache make software which is widely used and well known. In contrary to a proprietary software community, an OSS community is mostly worldwide distributed (143). In this paper, I focus on Online Design Discussion in OSS, especially on the work of Flore Barcellini, Francoise Detienne, Jean-Marie Burkhardt and Warren Sack in their published paper "A socio-cognitive analysis of online design discussion in an Open Source Software community". I will present their work by analyzing the used research strategy, the methods used to obtain their research data and their obtained results. Due to a worldwide distribution of developers in an OSS project, the community has to discuss design problems online via discussion spaces (mailing lists, forums, etc.) It is a distributed and asynchronous way of collaborative design work (143). This way of design work is less studied than other kinds of work designs like co-located collaborative design or distributed, synchronous design (143 / [De'04];[Ste02];[ea92]), that is why the work group has decided to study OSS project communities. In the focused paper they choose to study the Python project. This project has less than 100 developers involved, so the produced code is reasonable compared to other OSS projects like Linux, which contains hundreds of developers (146). It still has all the qualities common OSS projects have (146). Past studies from colleges in the workgroup also focused on Python. They had a solid base they could rely their work on. That is why they choose the Python project. They analyzed three aspects of online design discussions: Social dynamics, thematic temporal dynamics and design dynamics. Social dynamics represent the social influences participants have in the design process and how the status of a participant interacts in specific discussion and the chosen design strategy (142). The thematic temporal dynamics processes show how design themes emerge and evolve over time (142) and if they differ in their asynchronous way of discussion from face-to-face discussion themes (142). Design dynamics processes characterize how collaborative design work is done and how the community chooses to proceed with a design-issued proposal (142). They choose to analyze two design-related discussions within the Python-dev Python mailing list and then to compare and discuss the results.

2 Framework / Research Strategy

A major OSS project has usually five different types of participants (144): (1) The project leader, who is the creator of the project and approves final decisions. (2) The core team (administrators), they maintain the code and are involved in designing and evolving the software or some of its components. (3) Developers, who implement software components. (4) Active users, who participate in discussion space activity by reporting bugs or answering questions to newcomers. (5) Passive users who only use the software or read mailing lists and forums without any participation. [Mah02]

The Python project contained at the time of Mahendrans study ([Mah02]) a core group of seven people: The project leader himself and six administrators (146). The OSS design process of the Python project is separated into three design spaces. The documentation space (e.g. describes how an implementation of a design theme works and why it is done that way), the discussion space (most design-related issues are discussed here) and the implementation space (coding, etc.)



Distribution of the OSS design process: the case of Python.

The Python project has a certain protocol how to discuss and handle design related proposals made by a developer, administrator or the project leader himself. It is called "Python Enhancement Proposal" (PEP). The PEP protocols everything related to a proposal, including discussions, changes and the final implementation code. The participant who proposes a PEP is called champion. In order to understand the PEP process, I will give an example how a proposal will be handled:

- a) Champion sends PEP to an editor (editors are always administrators or developers)
- b) The editor checks the usability of the PEP. If confirmed, the PEP will be published and discussed within the discussion space.
- c) PEP will be discussed.
- d) PEP is approved by the community.
- e) Project leader gives his agreement.
- f) Code for PEP will be written and implemented into the Python language.

In order to maintain context within a distributed, asynchronous discussion community, certain quoting techniques are used by the participants. A quotation is used to identify a message as a response to another message (147). Quotation is widely distributed in the Python project discussion space and helps understanding and analyzing the structure of the community.

The work group focuses on the discussion space of the Python project, specifically the python-dev mailing list. They intended to develop a method that combines structural and content analyses of online discussions, based on quotation (147). They analyze three aspects of the Python project community: Social dynamics, thematic temporal dynamics and design dynamics (148).

2.1 Social dynamics

Social dynamics describes influences of participants in the design discussions, how they evolve during the design process and the relationship between the status of the posters and their involvement in the discussion (148). The work group analyzed four attributes to construct this relationship:

- a) Number of posted messages.
- b) Type of quotation structure.
- c) Depth of a quotation.
- d) Position in the discussion.

(148)

2.2 Thematic temporal dynamics

Thematic temporal dynamics show how design themes emerge and evolve over time (148). The work group analyzed design themes addressed by each message. This gives a characterization of the organization and the temporal emergence of themes in discussions (148). If a message is quoted, the degree of synchronism/asynchronism of this message will be calculated by the delay between the date the message has been first posted and the date the message has been quoted for the first time (148).

2.3 Design dynamics

Design dynamics characterizes how collaborative design work leads to a final design. Which sequences of activities and which discussion led to the final design? This is done by looking on messages containing design alternatives and analyzing them. For

example: Why is an alternative chosen or declined? In order to obtain such data, the work group had to perform a content-based study of the messages in the discussion space (148).

3 Methods

This chapter describes how the result data has been collected and analyzed.

3.1 Collection of the data

The PEP discussions are all archived and published on the Python-dev archives (<http://mail.python.org/mailman/listinfo/python-dev>). At the time of the study, 161 PEPs were discussed (148). The social structure of the Python project had been studied by colleagues of the work group. They gathered data for the first half of 2002 (148). In order to use this data for their social dynamics analyzes, the work group studied PEP processes from March and April of 2002 and assumed, that the Python project had the same membership during this period (148). They focused only on design-related PEPs to investigate the OSS design process (148). First they collected all PEP-related discussions by searching, by hand, the Python-dev mailing list for the keyword PEP in the subject header of each message (148-149). There were 342 different discussions, among them 53 discussions are dealt with 13 different PEPs (149). Eight PEPs were design-related PEPs. Among those PEPs, three have been just created and were chosen to capture the core of the design process (149). They compared two of these three PEPs chosen according to largest number of messages: PEP 279 and PEP 285. Both had the same status (accepted and implemented). According to F. Barcellini et al., they also were at the same stage in the design process and they addressed the introduction of new functionality. PEP 279 was proposed by a developer and PEP 285 by the project leader (149). So they differed in the social position of their respective champion (149). Also, PEP 285 had more meta-discussion issues like the backward compatibility of the new functionality (149); PEP 279 did not contain meta-related discussion (149). They compared these PEP processes by analyzing one main discussion (the one, were the champion proposes his PEP and asks for community input) for each of the two PEPs (149). This is a summary for the numbers the work group had to deal with:

PEP 279: 71 messages posted by 21 authors between March 28th and April 8th 2002; 3600 lines of text.

PEP 285: 95 messages posted by 22 authors between March 29th and April 5th 2002; 3500 lines of text.

3.2 Three analyzed aspects

Through a quotation-based representation of these PEP discussions they analyzed PEP's in the three aspects introduced before:

3.2.1 Social view

As explained above, the status of the participants had been analyzed by previous works their colleague Mahendran ([Mah02]) has done (149). Based on his work they identified four types of participants in the python-dev mailing list (149):

- a) Project leader (PL)
- b) Champion of the PEP (C)
- c) Administrators (A)
- d) Developers (D)

The involvement in the online discussion of each type of participant are characterized by a median number of messages posted (149). This categorizes the involvement in (1) high participant administrators (HP-A) / high participant developers (HP-D): Administrators (including the project leader) and developers (including the champion) who sent more than two messages (149). (2) Low participant administrators (LP-A) and low participant developers (LP-D) who posted two or fewer messages (149).

A line will be recognized as a quoted line, if it starts with a ">", for example:

xxx writes: >So now I'd like to choose between enumerate() and indexer(). Any >closing arguments? "indexer" is the name of the built-in full-text indexer, right? ;-) (149)

A message is named "text-only", if it does not contain any quotation; a single quote message (example above) is a message with one block of quotations followed by a comment (149); a multiple quotes message is a message "containing alternating quotes and comments from the same message or from different ones" (149). The work group investigated two aspects of the quotation patterns present in the discussions (150) by analyzing the messages position in the discussion graph and the depth of quotation (distance between source message and the last message that quotes it directly or indirectly) exhibited in the messages. For further details, please refer to 4.2 in this paper.

The figures in 4.2 represent a discussion graph and the depth of quotation for PEP 279. There are three different patterns of quotation distinguished:

- a) *Sequential structure: a message is quoted by only one message.*

- b) *Branching structure: a message is quoted by two or more messages.*
- c) *Closing structure: a message is in a closing structure when it is not quoted at all.*

(150)

3.2.2 Thematic temporal view

The work group identified a theme addressed for each message. In the PEP 279 discussion five themes have been identified and for the PEP 285 six themes (150).

PEP 279 themes:

- (T1): Choice of a name
- (T2): Alternative syntaxes for the proposed function
- (T3): Utility of the proposed function
- (T4): Error handling (how will the function inform the user of an error)
- (T5): Orthogonal problem of name binding / status of name spaces (technical issue)

PEP 285 themes:

- (T1): Consequences of a new build-in data type
- (T2): Specific function (return type and complications for existing Python code)
- (T3): Name of the constant of the new build-in type
- (T4): Elimination of non-boolean operations on booleans
- (T5): Specific operator of Python and its return value
- (T6): Inheritance relationship between two types (integer and boolean)

T1 of PEP 285 is a "meta-theme". The community discusses possible impacts this change could have on the whole Python programming language (150). The acceptance of this PEP is controversial, but it is still supported by the project leader. This makes it to a "meta-theme" (150).

They also analyzed the aspect of temporality within a theme-related discussion by extracting dates and server times from the headers of each message and measured the amount of time between the first posted message addressing a design theme TX and the final message(s) addressing TX.

3.2.3 Design view

The work group identifies design alternatives proposed and discussed in comments and the quotation patterns used by participants. A design alternative is a proposed solution by a developer to an addressed theme (151). For PEP 279/T1 23 design alternatives were proposed, eight for PEP 279/T2 and three for PEP 279/T3 (151). For the whole PEP 285's discussion there were only three alternatives proposed (151).

They also identified seven different design activities from the content of messages:

- a) Proposal: A developer proposes a new design theme / design alternative.
- b) Evaluation: A developer agrees or disagrees with a proposal.
- c) Group coordination: A developer moderates the discussion.
- d) Synthesis: The recapitulation of what has been discussed. In asynchronous design discussion a synthesis is used frequently to summarize the progress or process of a discussion.
- e) Clarification: A developer provides an explanation to address misunderstandings.
- f) Explicit decision: The project leader approves to a proposal. Only he can make explicit decisions.
- g) Other activities like humor or off-topic talk.

(151)

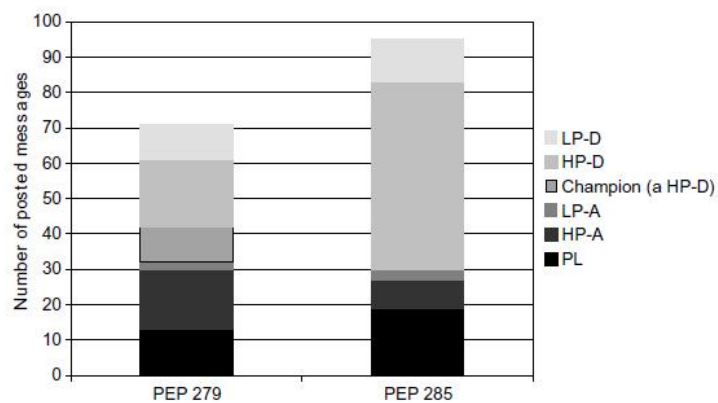
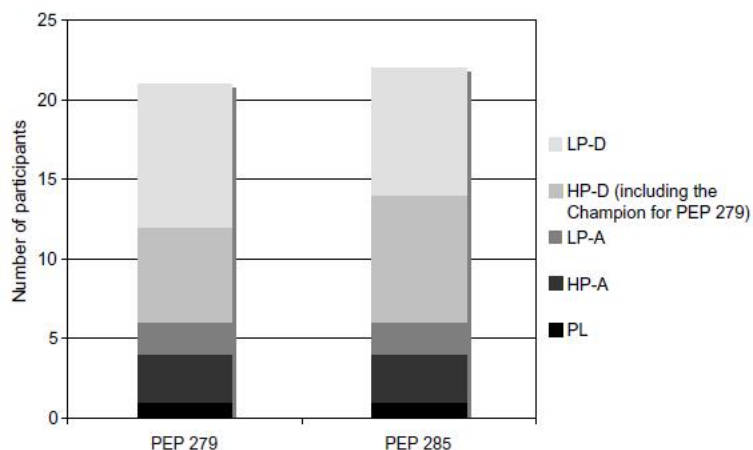
These seven activities are used in a coding scheme which assigns each message a content-based activity by searching them for words like "Why not ...?" (Proposal) or "yes" / "of course" / "great" (Evaluation) (151).

4 Results

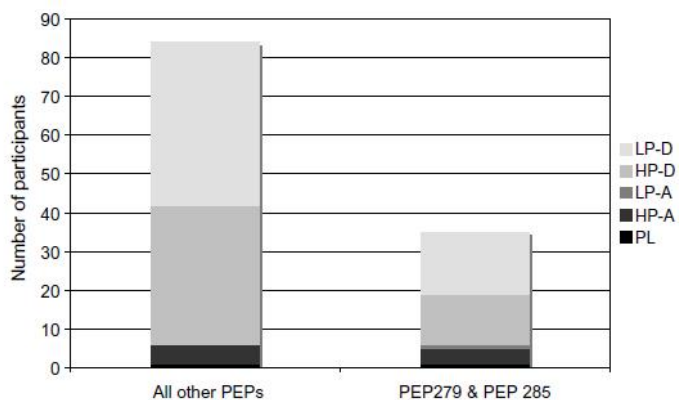
4.1 Social view

4.1.1 Global distribution of statuses and involvement in the discussions

Both discussions have similar distribution of participants according to their statuses coupled with their involvement (152):



Distribution of posted messages according to statuses and involvement of participants in PEP 279 and PEP 285.



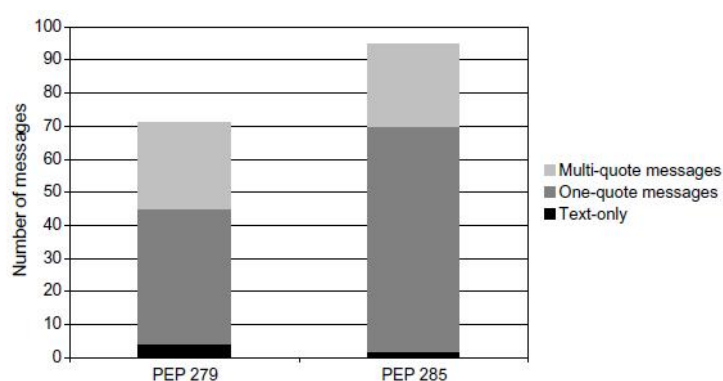
Distribution of participants according to their statuses and involvement in PEP 279 and 285 and in all other PEPs.

The project leader participated in both discussions frequently (between 18% and 20% of all messages). In PEP 279, the champion participated with 14% of the posted messages (152). This result is anticipated, they are expected to participate frequently in the

discussion (152). A local difference between the two PEP discussions has been found: More Administrators (HP-As) participated in PEP 279 discussions than in PEP 285 discussions (152). In contrast more developers participated in PEP 285 discussions than in PEP 279 discussions (152). This is the result of the importance of PEP 285: Changes considered in PEP 285 could have long-term implications of large numbers (152) for the Python community.

4.1.2 Quotation activities

F. Barcellini et al. verified, that "quotation is a common activity" (152). 94% (PEP 279) / 98% (PEP 285) (152) are quoted messages:

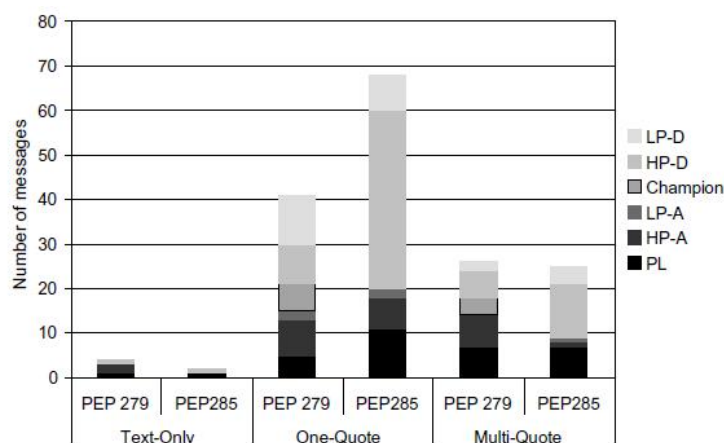


Distribution of quotations' structures in discussions for PEP 279 and PEP 285.

The quotation technique participants chose contained (selected) subparts of previous messages, which suggests, that the quotation is not determined by the default settings of the email clients the discussants used (153).

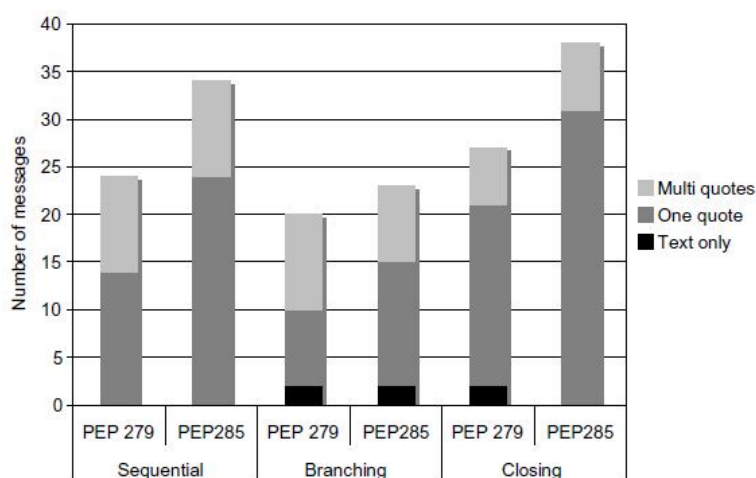
4.1.3 Position of participants and their used quotation structure

The PL intends to produce more multi-quote messages than other participants in PEP 279 (153), in PEP 285 multi-quote messages are used by the PL and by LP-As and LP-Ds:



Distribution of quotation structure according to status and involvement of participants for PEP 279 and PEP 285.

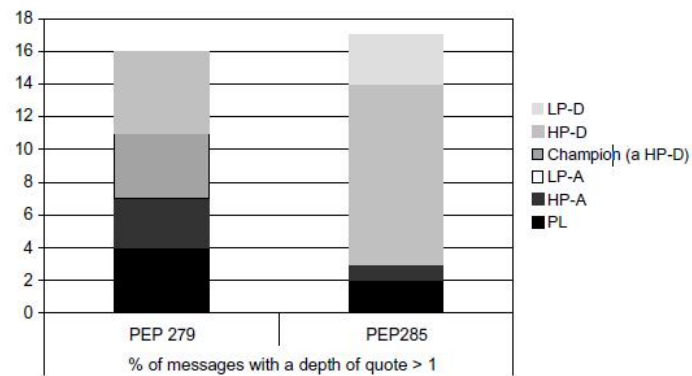
Further analyzes showed, that branching messages are frequently multi-quote messages than text-only messages (153) and closing messages are much more like to be one-quote messages. Text-only messages usually initiate discussions and first branching positions (153):



Distribution of quotation structures according to positions.

For PEP 279, branching structure is usually intended by a message posted by the PL or the C (153). Developer messages are mostly in sequential positions (153). Closing messages are initiated by the PL (153). For PEP 285, the PL has a big presence in the discussion and initiates branching structures (153). The presence of HP-As is very low in the discussion, they usually make closing messages (153). The most significant difference between PEP 285 and PEP 279 is that in PEP 285 the LP-Ds mostly initiate branching structures. This is the result of the importance of PEP 285 (153). There were previous debates about this PEP, so developers felt especially invested in the possible outcome (153).

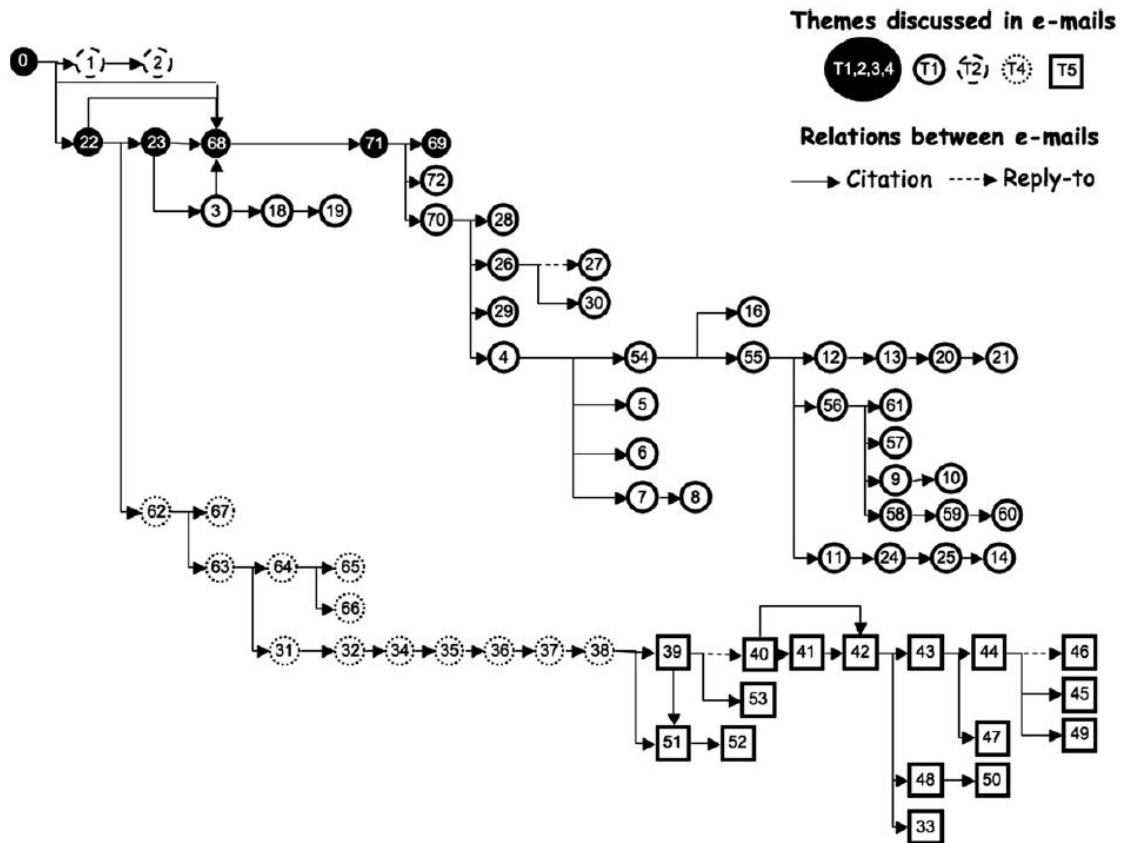
4.1.4 Depth of quotation

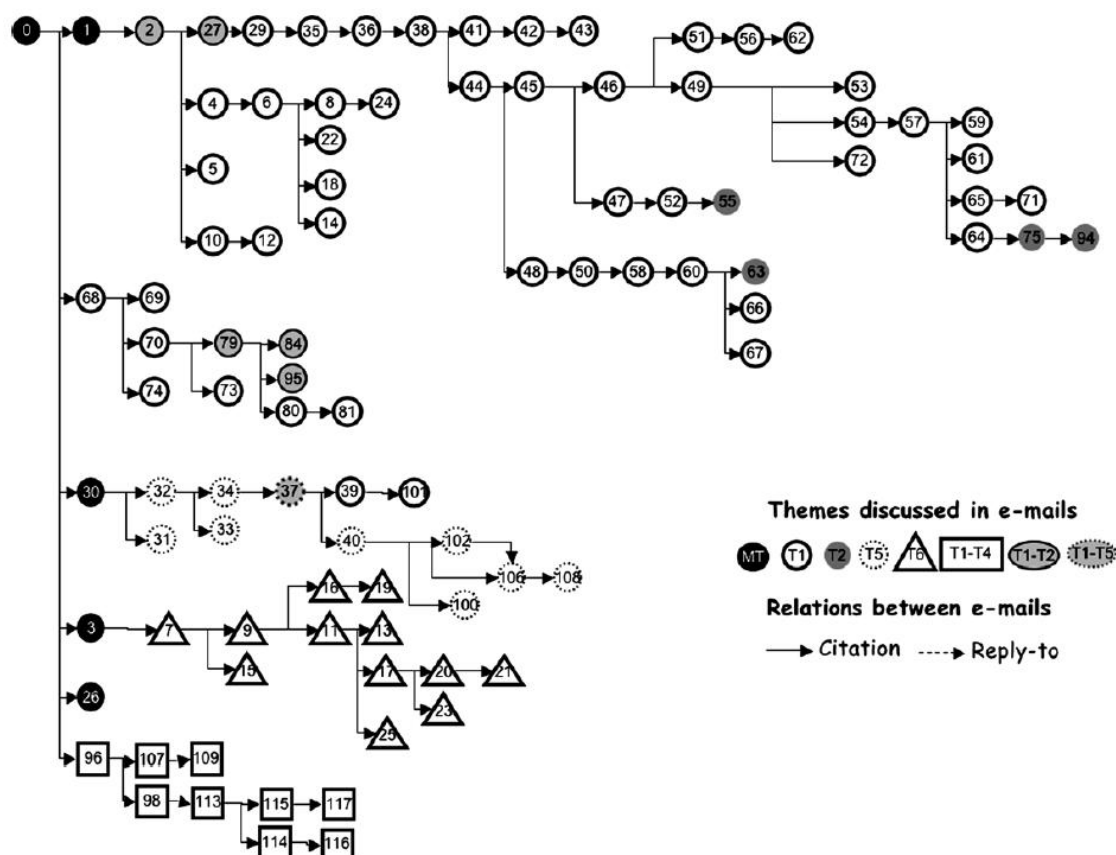


Distribution of messages with a depth of quotation >1 according to the status and involvement of the quoted participants for PEP 279 and PEP 285.

The work group revealed that deeply quoted messages in PEP 279 are more likely posted by the PL and the HP-As (155). As shown in the figure above, in discussions of PEP 285 HP-Ds and LP-Ds are more deeply quoted (155).

4.2 Thematic temporal view





Quotation-based representation of theme discussion themes (PEP 285).

The figures above show, how themes emerge during the discussions. First messages are usually multi-thematic posts and are mostly followed by single-theme messages (155). How long a theme is held up in a discussion depends upon the PL (155). The temporal degree of asynchronism and temporal coherence of discussions is characterized by the delay between the date the server has received a message and the date this message is quoted the first time (155):

Delay of appearance of 25%, 50%, and 75% of the 1st quotes for PEP 279 and PEP 285

Percentage of appeared quotes	PEP 279	PEP 285
25% of 1st quotes	20 min	15 min
50% of 1st quotes	1 h	2 h 16
75% of 1st quotes	5 h	7 h 33

As you see a message is usually quoted within 1h or 3h, which suggests a small degree of asynchronism (158) in the OSS community. Thematic temporal view also reveals that the PL only interferes in the discussion in the first few days and in the end of a discussion. In-between developers take over (158).

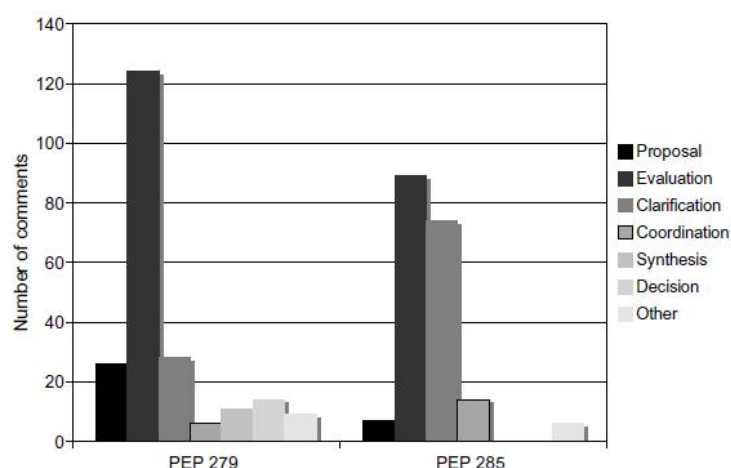
4.3 Design view

The results discussed in this section were gathered by content analysis of the messages. The work group analyzed design alternatives and design activities of the participants (158).

4.3.1 Design alternatives

For PEP 279, 34 design alternatives were proposed, in the PEP 285 discussion only three (158). The study reveals that branching structures result from messages containing more than one design alternative proposals; closing messages include one or zero design alternatives (158). In PEP 279 discussion alternatives were posted mostly by the C, the PL and LP-Ds (158); HP-Ds posted messages containing one or zero alternatives (158). The results show, how important the C and the PL are in the design process. PEP 285 contained fewer design alternatives because the discussion of design themes has been closed by the PL (159); developers had to choose between existing alternatives or to address other themes (159).

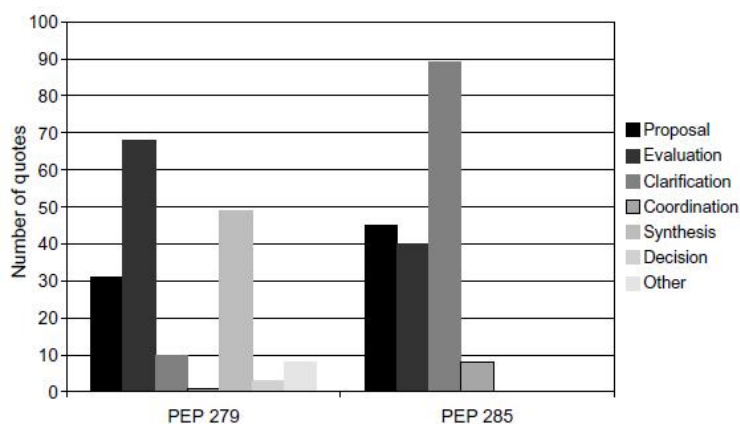
4.3.2 Design activities in comments



Distribution of activities in comments from the beginning portions of the PEP 279 and PEP 285 discussions.

Activities in comments show a difference between PEP 279 and PEP 285 discussions (158); the process of evaluation is in both discussions the most frequent activity (158). In PEP 285, clarification seems much more important than in PEP 279, which is reasonable considering the importance of the PEP 285 for the whole Python project community.

4.3.3 Design activities in quotes



Distribution of activities in quotes PEP 279 and PEP 285.

As presented on the figure above, most of the activities quoted differ between PEP 279 and PEP 285 (159): The most quoted activities in PEP 279 are synthesis and evaluation; in PEP 285 clarification is the most quoted activity (159).

4.3.4 Sequences of activities

The study reveals, in which sequences activities appear online (159). They characterize the way information and knowledge is gathered in the Python project OSS community (159).

- a) An evaluation is usually followed by another evaluation (at example: positive agreement is followed by a negative agreement)
- b) A proposal is usually followed by an evaluation or other proposals.
- c) A clarification is usually followed by a proposal or agreements.
- d) A coordination activity is usually followed by another coordination activity.
- e) A decision is usually followed by a coordination activity (corresponding to task assignments implied by the decision).

(159)

5 Discussion

This section discusses the main results the work group gained with its study. The results and the methodology used to obtain the analyzed data can be used to implement and automate the gathering process, so more data can be analyzed. The work group did that mostly "by hand".

5.1 Social dynamics

The PL and C of the PEP seem to be in all discussions frequent contributors (159). Messages posted by them tend to lead to multiple branches (159). Messages of the PL usually contain multi-quotes (160). He also makes the final decisions and close discussions (160). The C is the only participant writing synthesis to gather the information of previous posts (160). The study reveals the absolute authority the PL has in the project (160). Future work may investigate how the C of the PEP affects the design of the project. Messages posted by administrators occurred in three different positions in the discussion (160):

1. At a beginning branching position.
2. In linear exchanges with developers.
3. In closing positions (acting as the PL when he was not available).

(160)

In the PEP 285 administrators participated only in meta-theme discussions (160) (due to the fact, that design-alternative discussion has been closed by the PL). The position c) shows, that administrators can take other roles in the community and may fill in for the PL if he is not available (160). The PL acts when a crucial theme for the Python language appears and the administrators seem to rely completely on him (if he is present).

Messages posted by developers occurred in two different positions in the discussion (160):

- a) In PEP 285 at a beginning branching position.
- b) In PEP 279 in linear sequences of exchanges with other participants.

(160)

This shows that developers post design alternatives and are actively participating in the design process of the project.

5.2 Thematic temporal dynamics

The quotation technique used by participants is based on selective quoting (160). They choose the lines necessary to preserve context in their own messages and to avoid misunderstanding (160). The timeframe participants react to a proposal or a message is different. If a message is considered important (160) the answer usually appears within hours. In other non-important design-discussions an answer is not estimated "in time" (160). There are certain moments in a design-oriented, distributed online discussion where the communication can be seen as synchronous (160) and it seems that implicit social rules exist in an OSS project community; if somebody does not answer an important issue "in time", he could lose his social status within the community (160).

5.3 Design dynamics

5.3.1 Activities

Similar to face-to-face design meetings (Olson et al., 1992; d’Astous et al., 2004) in distant and asynchronous discussions the two most important activities are evaluation and clarification (161). In difference to face-to-face design meetings were many decisions are made in an implicit manner, in online discussions much fewer decisions are taken, but those taken are very explicit (161). A participant’s status does usually not seem to affect the design process he is engaged in with the exception that certain activities are tied to a certain status, for example only the PL himself approves / disapproves of a proposal and makes the final decision if to proceed with a solution (161).

5.3.2 Sequence of activities

The patterns of interchange do not differ in face-to-face discussions (161). For example: An evaluation comment tends to be followed by another evaluation comment, etc.; that does not differ in face-to-face discussions (161).

6 Summary / Future

As you see, this paper is a summary of the work done by F. Barcellinis work group. I intended to give a short overview how design-related discussion in an OSS project community works. In the future, the work group is planning to analyze other types of PEP discussions like meta-PEPs, rejected PEPs, still open PEPs and PEPs with different champions in the Python project to understand how participants interact within a community and to provide a basis for a social study of online discussion spaces.

References

- [De´04] Boujut-J.-F. Hohmann-B. De´tienne, F. *Characterization of collaborative design and interaction management activities in a distant engineering design situation.*, 2004.
- [ea92] Olson et al. *Small group design meetings: an analysis of collaboration.*, 1992.
- [FB08] Jean-Marie Burkhardt Warren Sack Flore Barcellini, Franc,oise De´tienne a. *A socio-cognitive analysis of online design discussions in an Open Source Software community*, 2008.
- [Mah02] D. Mahendran. *Serpents and primitives: an ethnographic excursion into an Open Source community.*, 2002.
- [Ste02] Badke-Schaub P. Stempfle, J. *Thinking in design teams: An analysis of team communication.*, 2002.