EE382V – Software System Measurement & Metrics - Assignment 3

Deadline: Friday, November 11, **before** midnight (late reports will not be accepted!!!)
Grade: 20 points - each team member receives the same initial score, then adjusted by their individual participation level.
Submittal process: submit your team's report online using the assignment manager feature on the class blackboard page. *Don't forget to put all team member names on it.*

Purpose. Perform a retrospective, data-driven case study of an open source repository of useful software. The subject of our case study is an open-source project called Django. It's a community development of a Python-based Web framework. The project's website is found at:
https://www.djangoproject.com/

Potential Sources of case study data:

You can browse the code base online at:
https://code.djangoproject.com/browser/django/trunk
You'll need Subversion if you want to download the code base.

On obtaining the source code, here's a good document that gives an overview of Django's subversion repository and the the ways to access it:
https://docs.djangoproject.com/en/dev/internals/svn/

Here's a document that describes how to run Django's unit tests:
https://docs.djangoproject.com/en/dev/internals/contributing/writing-code/unit-tests/

There is also an automated build server that runs the unit tests against several configurations
http://ci.djangoproject.com/job/Django/

Django's ticket (e.g.bug) database lives here:
https://code.djangoproject.com/query
You can download these to a spreadsheet for further analysis.

Getting at data dumps and APIs for Django's bug and commit data (in Trac):
https://code.djangoproject.com/wiki/OpenData

As for Python static analysis tools, you could use PyMetrics:
http://sourceforge.net/projects/pymetrics/

You could do current report on the cyclomatic complexity measures of the Django code base, like Gary did over 5 years ago. Post a new version of that report.
http://gdub.wordpress.com/2006/07/09/cyclomatic-complexity-of-django/

There's a code coverage tool, called coverage.py, that can do both line and branch coverage metrics:
http://nedbatchelder.com/code/coverage/

View and analyze the detailed timeline of project events:
https://code.djangoproject.com/wiki/TracTimeline

https://code.djangoproject.com/timeline

Discovery and use of other Python measurement tools is encouraged.

Case Study Assignment

A case study is an empirical method, i.e. a defined, scientific method for posing research questions, collecting data, analyzing the data, and presenting the results. Each of these steps is planned from the outset of the study. Case studies are well suited to "how" and "why" questions in settings where the researcher does not have control over variables and there is a focus on contemporary events. Case studies are used primarily for exploratory investigations, both prospectively and retrospectively, that attempt to understand and explain phenomenon or construct a theory. They are generally observational or descriptive in nature. Due to this dexterity, they have become quite popular in software engineering and are frequently used in papers to understand, to explain or to demonstrate the capabilities of a new technique, method, tool, process, practice, technology or organizational approach. The six steps of a case study are:

1. Determine and define the research questions
2. Select the case object of study and determine data gathering and analysis techniques
3. Prepare to collect the data
4. Collect data
5. Evaluate and analyze the data
6. Prepare the report

Step 1. Determine and Define the Research Questions
The first step in a case study is to establish a firm research focus to which the researcher can refer over the course of study of a complex phenomenon or object. This is done by forming questions about the situation or problem to be studied and determining a purpose for the study. The research object in a case study is often a project, an entity, a person, or a group of people. To assist in targeting and formulating the questions, researchers will conduct a focused literature review to see what relevant research has been previously conducted. This often leads to refined, more insightful questions about the problem. Careful definition of the questions at the start pinpoints where to look for data/evidence and helps determine the methods of analysis to be used in the study. The literature review, definition of the purpose of the case study, and early determination of the potential audience for the final report guide how the study will be designed, conducted, and reported.

The study questions of interest are, at least:

1. **What is the size of the code base? And how fast is it growing?**
2. **What is the cyclomatic complexity of all the modules? List all the modules that should be refactored or redesigned to simplify them.**
3. **What is the quality of the various components, based on collected defect data (reported as a Pareto diagram)? What was the trend leading up to the most recent release ?**
4. **Other interesting questions that I have not thought of yet.**

Step 2. Select the Study Object & Determine Data Gathering/Analysis Techniques
During the design phase of the case study, the researcher determines what approaches to use in selecting the case to examine in depth and which instruments and data gathering approaches to use. Careful discrimination at the point of selection also helps erect boundaries around the case. A useful step in the selection process is to repeatedly refer back to the purpose of the study in order to focus attention on where to look for evidence that will satisfy the purpose of the study and answer the research questions posed. The researcher determines in advance what evidence to gather and what analysis techniques to use with the data to answer the research questions. Data gathered is normally largely qualitative, but it may also be quantitative. Tools to collect data can include surveys, interviews, documentation review, observation, and even the collection of physical artifacts. The researcher uses the designated data gathering tools systematically and properly in collecting the evidence. Throughout the design phase, researchers should ensure that the study is well constructed to ensure construct validity, internal validity, external validity, and reliability. **The subject of this semester's study is the Django**

repository of information (and any other data about it that you discover).

Step 3. Prepare to Collect the Data
Advance preparation assists in handling large amounts of data in a systematic fashion. For example, researchers prepare databases to assist with categorizing, sorting, storing, and retrieving data for analysis. Researchers must also be aware that they are going into the world of real human beings who may be threatened or unsure of what the case study will bring. Researchers need to anticipate key problems and events, identify key people, prepare introductions, establish rules for confidentiality, and actively seek opportunities to revisit and revise the case study design in order to address and add to the original set of research questions as needed.

4. Collect Data in the Field
The researcher must collect and store sources of evidence comprehensively and systematically, in formats that can be referenced and sorted so that converging lines of inquiry and patterns can be uncovered. Researchers carefully observe the object of the case study and identify causal factors associated with the observed phenomenon. Renegotiation of arrangements with the objects of the study or addition of questions to interviews may be necessary as the study progresses. Case study research is flexible, but when changes are made, they are to be documented systematically. Exemplary case studies use field notes and databases to categorize and reference data so that it is readily available for subsequent reinterpretation. Field notes record feelings and intuitive hunches, pose questions, and document the work in progress. They record testimonies, stories, and illustrations that can be used in later reports. They may warn of impending bias because of the detailed exposure of the subject to special attention, or give an early signal that a pattern is emerging. They assist in determining whether or not the inquiry needs to be reformulated or redefined based on what is being observed. Field notes should be kept separate from the data being collected and stored for analysis. The researcher documents, classifies, and cross-references all evidence so that it can be efficiently recalled for sorting and examination over the course of the study.

Step 5. Evaluate and Analyze the Data
The researcher examines raw data using many interpretations in order to find linkages between the research object and the outcomes with reference to the original research questions. Throughout the evaluation and analysis process, the researcher remains open to new opportunities and insights. The case study method, with its use of multiple data collection methods and analysis techniques, provides researchers with opportunities to triangulate data in order to strengthen the research findings and conclusions.

The tactics used in analysis force researchers to move beyond initial impressions to improve the likelihood of accurate and reliable findings. Exemplary case studies will deliberately sort the data in many different ways to expose or create new insights and will deliberately look for conflicting data to disconfirm the analysis. Researchers categorize, tabulate, and recombine data to address the initial propositions or purpose of the study, and conduct crosschecks of facts and discrepancies in accounts. Focused, short, repeat interviews may be necessary to gather additional data to verify key observations or check a fact.

Specific techniques include placing information into arrays, creating matrices of categories, creating flow charts or other displays, and tabulating frequency of events. Researchers use the quantitative data that has been collected to corroborate and support the qualitative data that is most useful for understanding the rationale or theory underlying relationships. When the multiple observations converge, confidence in the findings increases. Conflicting perceptions, on the other hand, cause the researchers to pry more deeply. In all cases, the researcher treats the evidence fairly to produce analytic conclusions answering the original "how" and "why" research questions.

Step 6. Prepare the report
Exemplary case studies report the data in a way that transforms a complex issue into one that can be understood, allowing the reader to question and examine the study and reach an understanding independent of the researcher. The goal of the written report is to portray a complex problem in a way that conveys a vicarious experience to the reader. Case studies present data in very publicly accessible ways and may lead the reader to apply the

experience in his or her own real-life situation. Researchers pay particular attention to displaying sufficient evidence to gain the reader's confidence that all avenues have been explored, clearly communicating the boundaries of the case, and giving special attention to conflicting propositions. Some researchers report the case study as a story. Some case study researchers suggest that the documents should be reviewed by the participants in the study as a reality check.