Data Sharing, Distribution and Updating Using Social Coding Community Github and \LaTeX Packages in Graduate Research

Presented at
Integration of the Education and Science: Challenges of the Modern World
Kazakh-Russian International University
Aktobe, Kazakhstan

Polina Lemenkova
pauline dot lemenkova at gmail dot com

May 13, 2015
Outline

1. Introduction
   - GitHub. What Does it Mean?
   - GitHub: Shared Creativity
   - GitHub: Creative Sharing

2. Advantages of Web-Service GitHub
   - \LaTeX{}: Advantages over Traditional Text Editors

3. How to Create Personal Repository?
   - Step 1: Key Generation
   - Step 2: User Registration
   - Step 3: Generating Project
   - Step 4: Maintaining Project Push and Update

4. Typesetting Thesis in \LaTeX{}

5. Why Using GitHub?
   - Functionality of GitHub
   - Advantages of GitHub in Academia
   - System of Control and Revision
   - Maintained Privacy
   - Online Feature of GitHub: Data-Sharing

6. A Case Study of Using GitHub
   - Project Interface in the Github Environment
   - Example of Research Progress
   - Adding New Data to the Project
   - Access to the Project

7. Popularization

8. Conclusion

9. Discussion

10. Résumé

11. Thanks

12. Bibliography

Polina Lemenkova

Data Sharing, Distribution and Updating Using ... Github and \LaTeX{}
Some Facts

- Octocat (Octopus + Cat) is a symbol of GitHub. It symbolizes sharing a project.
- GitHub is available for free which is useful for students and researchers.
- Slogan of GitHub: ‘Social Coding’, i.e. ‘let’s code it together’.
- It refers to the programming, but in this presentation I show using GitHub for MSc thesis.

- GitHub - a web-service for hosting (i.e. serving and maintenance) of IT-projects online, as well as their development by multiple authors (incl. graduate projects development).
- GitHub is initially aimed at version control system Git (used in programming).
- GitHub is written in the programming language "Erlang" and "Ruby", a framework Ruby on Rails.
- GitHub was created by GitHub, Inc, USA. The first private repository was established in 2008.
GitHub allows viewing and editing texts: make edits, change current versions in collective access; work together on the current version of the project; add new colleagues for collective work (convenient, for example, when a group of students write a common fieldwork report). GitHub maintains colored syntax highlighting: added parts of code/text are colored green, deleted parts of code/text – red. Example: below (fragment of my MSc thesis code).
GitHub: Creative Sharing

- Github allows adding various data types: code, graphics, etc.
- Github allows control latest changes, discuss and discuss work with students, post any comments directly into the text, add online comments
- Github allows registered users to add current version changes to the repository of the current project.
- Github keeps all update versions in the current projects online
- Github gives the opportunity for data sharing
- Github enables coworking
Advantages of Web-Service GitHub

- Online repositories
- Possibilities for co-authorship
- Creating PhD/MSc/BSc theses
- Presenting research works
- Writing and updating articles
- Archive of scientific works
- Sharing with colleagues
- Notification of current updates
There are a number of drawbacks when using traditional programs for writing theses (e.g., MS Word). These difficulties are essential when writing a thesis:

- Continuous numbering of the test and illustrations breaks when changing or adding new ones in the middle of the text (which must be corrected manually)
- In case of MS Word, student has to double-check all references to literary sources again, which complicates the work and leads to mechanical difficulties and lengthy monotonous corrections.
- \LaTeX{} has a built-in flexible system of bibliographic cross-referencing in the list of references, which enables making automatic linking to the bib sources, as well as updating links.
- \LaTeX{} has a built-in BibTeX package that enables compiling bibliography in active mode and leaves hyper refs a d live links to references, i.e. Instantly editable if necessary.
How to Create Personal Repository?

Step 1: Key Generation

```
please make sure you have the correct access rights
and the repository exists.
$ polina-mbp:TeX_CV pauline$ cat ~/.gitconfig
[User]  
  email = pauline.lemenkova@gmail.com
  name = pauline.lemenkova
$ polina-mbp:TeX_CV pauline$ git config --global user.name "Pauline Lemenkova"
$ polina-mbp:TeX_CV pauline$ ssh-keygen
generating public/private rsa key pair.
inter file in which to save the key (/Users/pauline/.ssh/id_rsa):
inter passphrase (empty for no passphrase):
inter same passphrase again:
our identification has been saved in /Users/pauline/.ssh/id_rsa.
our public key has been saved in /Users/pauline/.ssh/id_rsa.pub.
the key fingerprint is:
polina@polina-mbp
he key's randomart image is:
```
Step 2: User Registration

git config - global user.email 'you@example.com'
git config - global user.name ‘Your Name’
set up user’s account default identity
Step 3: Generating Project

Key Commands and Tools:

- `git init` - initiation project from scratch.
- `git add files` - selecting all files for the project. (texts, tables, graphs, maps, figures).
How to Create Personal Repository?

Step 4: Maintaining Project Push and Update

Project Push and Update

- `git commit -a -m 'name of update'`
- e.g.: `git commit -a -m 'added tables No 14, 15, 17'`
- 'git diff' - a key command of GitHub,
- 'git diff' - detects and recognizes all updates in the text
- 'git diff' highlights them green/red, respectively

```
commit 33236a417e284fc6014438dcd90c18372d5f006d
Author: paulinelemenkova <pauline.lenemankova@gmail.com>
Date: Sat Dec 20 18:27:06 2014 +0200

   update 2014

commit 380322d6c42d7b8117b22d17d2b7ad7f4b1b0f
Author: Paulina Lemenkova <pauline@hve-0981.koleje.cuni.cz>
Date: Fri Jan 31 15:33:16 2014 +0100

   My CV update as of 31.01.2014

commit d450df8134375676ff6fc6d3686f1922cfb512
Author: Pauline Lemenkova <pauline@para.gt.rtu.edu.tw>
Date: Thu Apr 11 14:50:53 2013 +0300

   updated my CV, Taiwan

commit e3b04e2335d5e47e335679300e1a00d6f02757
Author: Pauline Lemenkova <pauline@Pauline-Lemenkova-MacBook-Pro.local>
Date: Wed Oct 24 17:36:37 2012 +0200

   current update of CV

polinas-mpb:TeX_CV pauline$ git show
commit 33236a417e284fc6014438dcd90c18372d5f006d
Author: paulinelemenkova <pauline.lenemankova@gmail.com>
Date: Sat Dec 20 18:27:06 2014 +0200

   update 2014

diff --git a/CV_Lemenkova.tex b/CV_Lemenkova.tex
index 0f62b0...fa3d012 100644
--- a/CV_Lemenkova.tex
+++ b/CV_Lemenkova.tex
@@ -1,10 +1,10 @@

\documentclass[10pt]{article}
\usepackage[calc]
-\usepackage[english]{babel}
\usepackage[english]{babel}
```

Polina Lemenkova

Data Sharing, Distribution and Updating Using ... Github and \LaTeX

May 13, 2015
Typesetting Thesis in LaTeX

Example of structuring text in a thesis with the help of mark up language used to highlight text when writing codes with a high level of nesting, allows you to quickly navigate the work: paragraphs, highlighting text, indentation, multi-level markings, making tabs, hierarchical structuring (chapters, sections, sub-sections, paragraphs), multi-level indents from the red line. This allows to quickly read and navigate within the text.
Why Using GitHub?

Syntax Coloring

- Important feature of Github: built-in color management for updated code
- All recent updates are syntactically highlighted green in the command line

- Added text, lines and whole fragments and paragraphs of paragraphs are colored green
- On the contrary, selected sections of text deleted from the last session are colored red.
Advantages of GitHub in Academia

Time Monitoring

Using GitHub facilitates monitoring research progress and to assess the work done recently, to timely response to the comments and corrections of colleagues.

Retrospective Editing

Retrospective and comparative editing of texts in the Github environment allows to return to the old, previous version of the work, saved a while ago, to cancel updates.
System of Control and Revision

All versions and changes of the thesis are recorded and available in the system. Thus, using GitHub revision control system and text code management, the project was regularly updated. Hence, supervisors can regularly monitor the project.
Unpublished Works

Both public and private repositories are maintained: the colleague/co-author/supervisors may receive private links to the current version of research and have access to the work.
Why Using GitHub?

Online Feature of GitHub: Data-Sharing

Github’s use of the standard command line interface in programming

- Adding current version changes to the repository of the current project.
- Supports both a public repository and a closed one in limited access (private).
- Keeping closed unpublished works, with restricted access to coworkers.
- Successfully defended dissertations, master’s projects and theses can be posted in open access.
- Access is configured for both the general public and the university archive.
- Data can be stored there for an unlimited time.
A case study of using GitHub: my MSc thesis.

- MSc Thesis “Seagrass Mapping and Monitoring Along the Coasts of Crete, Greece”, defended in Netherlands, University of Twente, Faculty of GIS and Earth Observation, 2011.
- Written using text editor \texttt{\LaTeX} using GitHub
- Original source: on my GitHub webpage.
## Project Interface in the Github Environment

### Opportunities
GitHub: excellent opportunities in academia for supervisors, students, lecturers, researchers.

### Monitoring
The use of \LaTeX\ and Github provided timely access for the project’s supervisors to monitor research progress.

### Examples
Updating Unlike traditional MS Word, combination of \LaTeX\ and Github allowed supervisors to check current progress on-line as updates, comments and corrections made.
A Case Study of Using GitHub

Example of Research Progress

New text is highlighted in green, deleted paragraphs are highlighted in red.

Current research progress can be easily monitored and specific project updates highlighted and commented.
Adding New Data to the Project
Access to the Project

This section likely discusses methods and tools for accessing and managing project-related data, potentially using GitHub. It might cover aspects such as collaboration, version control, and access control, which are critical for effective project management. The context suggests a focus on integrating GitHub with other tools like LaTeX, possibly for scientific or academic projects, where collaboration and document management are essential.
Actuality of Using GitHub

Audience

Despite obvious advantages and prospects of the GitHub service, the majority of its users are programmers and IT specialists. Using GitHub in academia is still limited.

Popularization in Academia

There is a need to popularize and demonstrate the GitHub service and environment in the student and academic environment: in research centers, universities, institutes.

Examples

Geologists: Getting acquainted with GitHub would be of especial advantage for faculties of natural and Earth sciences, since they often have common fieldwork data and projects.
Conclusion

Current problem with GitHub and \LaTeX{} is their non-popularity in academia caused by following reasons.

**Student Works**
Some students are not informed about the \LaTeX{} or GitHub and do not know how to use them. The functionality of these tools, multifunctional environments should be popularized.

**Difficulty of Learning Curve**
GitHub and \LaTeX{} are sometimes regarded as difficult programs. The basics of using \LaTeX{} and GitHub should be initially studied for some time, but it is rewarding. Further advantages of using \TeX{} and GitHub are obvious and worth the efforts and time to master them.

**Examples**
Advantages: Joint combination of using \LaTeX{} and GitHub facilitates writing, typesetting and managing versions upgrades while working on a project.
## Discussion

### Opportunities

Traditional for IT industry and programmers, GitHub offers great opportunities for effective collaboration such as 'student-supervisor' or 'group of students' or 'group of colleagues'. Unlike traditional programs, the use of innovative technologies of text editors, such as \( \LaTeX \) and data archiving such as GitHub allows supervisor to monitor student’s research progress.

### Advantages

A great advantage of GitHub consists in its color syntax and maintaining history of updates versions. The presentation demonstrated conceptual principles of the Github and \( \LaTeX \).

### Possibilities

The possibilities for collaborative data sharing, research progress updating and creative works are illustrated. Current work is a technical illustration of using IT in education.
This work aimed to give an example of the effective use of the GitHub web service specifically in an academic environment. The case study was given to a MSc thesis that was written completely using \LaTeX and GitHub. The presentation illustrated how specifically one can apply new technologies and innovative approaches in the educational environment.

Technical issues of running project, creating repositories, making updates, adding new files to the system are demonstrated with a series of screenshot illustrations of the process. Important conceptual features and advantages of using Github and \LaTeX in the academic environment are listed and discussed.
Thank you for attention!


24. P. Lemenkova, “Geoecological Mapping of the Barents and Pechora Seas”, Russian, B.Sc. Thesis (Lomonosov Moscow State University, Faculty of Geography, Department of Cartography and Geoinformatics, Moscow, Russia, May 18, 2004), 78 pp., https://thesiscommons.org/bwcr.


