A “good” description on GitHub :

Language x Software Technology x Purpose (LSP)

Automatically Categorizing the Purpose of GitHub Repositories

Jazlyn Hellman; B.A. Software Engineering & IDS
Jin Guo; McGill University **Christoph Treude**; University of Adelaide

INTRO

There are millions of repositories on GitHub designed for sharing, collaboration, and discovery, but there is a lack of requirements or standardized ways to present information**[1]**. As a result, many repositories' descriptions are unclear, have outdated information, or empty making the search process tedious and difficult.

Objective

From a repo's URL, automatically generate an informative, understandable description of the project.

Procedures

1. GitHub RepoReaper Data**[2]**
2. Descriptions➡Readmes➡Wikipedia**[3]**
3. **LSP Template & Labelling**
4. Gather & Analyze Dependency Trees**[4]**

RESULTS

- Most repositories' initial descriptions did not fulfill the LSP Template and were not informative.
- Natural Language Processing Techniques are likely to identify the Purpose.

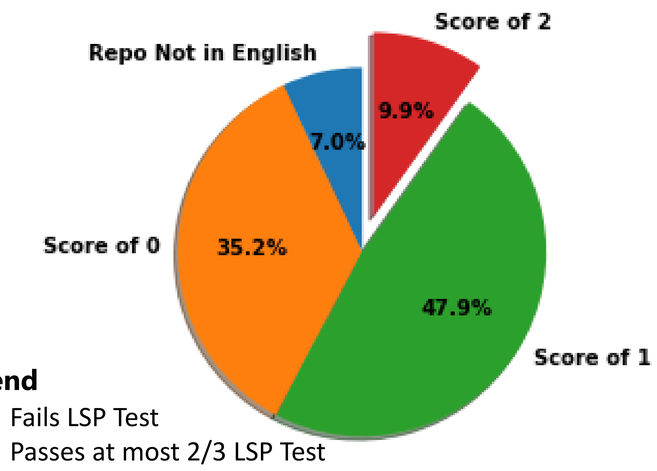
Future Work

- Complete dataset creation
- Create a Classifier for existing descriptions based on LSP
- Automatically identify Purpose
- Automatically generate 'good' descriptions from a URL
- Conduct evaluations and validate LSP template

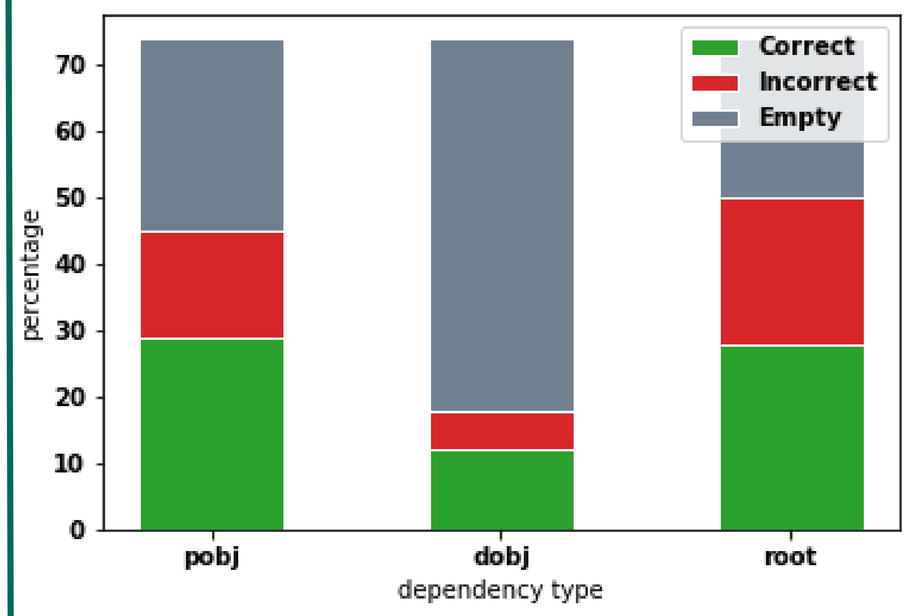
I want to express my thanks and gratitude to Mr. Harry Samuel and the ARIA program for funding my research.

References
[1] https://github.com
[2] Munaiah, N., Kroh, S., Cabrey, C. et al. Empir Software Eng (2017) 22: 3219. https://doi.org/10.1007/s10664-017-9512-6
[3] Brank, J., Leban, G., & Grobelnik, M. (2017). Annotating documents with relevant Wikipedia concepts.
[4] https://github.com/explosion/spaCy

Procedure 3: LSP Scoring



Procedure 4: Dependency Type as a Purpose Indicator



Procedure 2: Sample LSP-Descriptions

Legend
- Language
- Software Technology
- Purpose
- Delivery Format

Repository Name	Initial Description	LSP-Description
AgilTec/cadenero	Rails.API Authentication Engine for multitenant RESTful APIs	Ruby implementation of a Ruby on Rails API Authentication Engine for consumers of multitenant RESTful APIs services
brenoc/opentracks	A Flask/Python client for Open-Transactions.	Python implemented Flask microweb client for the open-source financial technology Open-Transactions project.
onaio/onadata	Collect, Analyze and Share	Python implemented platform for data collection/analyzation/ sharing that utilizes the Sphinx and Django frameworks.
clementine-player/ Android-Remote	Control Clementine from your Android device	Java powered android application to remotely control your open-source music player Clementine.
rafallo/p2c	New way to watch video!	Python powered desktop application for streaming pre-downloaded content through the BitTorrent communication protocol.