Software Language Processing

(IR++)
Max, Nasir, Chanchal, Latifa, Wei, Fabian, Laura, Mario, Hongyu, Jim, Rafael, Tom, Tao, Peter, Stephe, Baishakhi, Foyzur, Akinori, Peter, Joshua, Prem, Nic, Abram

How can we help developers?

- Traceability
  - Email -> Source Code Traceability
  - Bug -> Source Code Traceability
- Bug de-duplicating
- Awareness (topic timeseries)
- Code Completion
- Triage
- Bug prediction
- IDE querying
- Summarization (Code/Bugs)
- Test Case Prioritization

Challenges in SLP/IR++

- Context Modelling
  - Project/Domain/SE Context
- Sound Methodology for applying SLP/IR++
  - Preprocessing and methodology (preprocessing, splitting) (Stop word lists and dictionaries)
  - Appropriate Parameters per task
- Finding the limitations with respect to software and SE artifacts
  - Implicit semantics
- Exploiting IR effectively
  - Understand the assumptions
  - How to transform SE data appropriately for IR
- Lack of availability of markedup data.

Breakout Activities

- Markup/Discuss a bug report (6)
- Markup/Discuss an Email Thread (5)
- Markup/Discuss a function or some source code (8)
- IR Task Walkthrough (by hand?)
- Your suggestion....