

IR++:

1. IR
2. NLP
3. Contextual
4. Ontology
5. Code Analysis
6. MSR + SLP
7. Modelling

Why?

1. Lightweight Techniques
2. Ease of use
3. Some NL similarity
4. More skewed distributions

Problems

1. Traceability (Email -> Src)
2. Stakeholder Language
3. IR inappropriateness (Lack of SE knowledge)
4. Leverage for bug fixing
5. Informal Documentation -----> Alignment (in technique section)
6. Bug -> Src
7. IR Parameters
8. Identifier grammars
9. Partial Code

Techniques:

1. Code -> Text Context
2. TF-IDF
3. NLP
4. Alignment
5. LDA, LSI, etc.
6. NLP Grammars
7. Segmentation
8. Language Models
9. Corpus+Context (API+Security Policy)

Activity Idea:

1. Markup and bug report
2. Markup function
3. Challenge list
4. Email

Challenges:

1. Context modelling