Source Code Exploration with Google

Denys Poshivanyk, Maksym Petrenko, Andrian Marcus, Xinrong Xie, Dapeng Liu

SEVERE Group @ Wayne State University
Issues in concept location

• Simple and efficient methods -> poor precision and recall

• Complex methods -> higher precision and recall at the cost of efficiency
IR based concept location

• Uses Latent Semantic Indexing to index and search the source code
• Problems with the prototype
  - slow
  - inefficient re-indexing
  - no resources to support and adoptable version
Solutions for adoption

• Industry grants
  - not always fits our research agenda

• Component based approach with industrial strength COTS
Remember WCRE2004

Google™
Your Software System

Het Internet Afbeeldingen Discussiegroepen Gids
Google zoeken Ik doe een gok
Het web doorzoeken Zoeken in pagina’s in het Nederlands

Geavanceerd zoeken Voorkeuren Taalhulp middelen

ICSM 2006
Google Eclipse Search (GES)
GES interface
GES results
Advantages

• Support of natural language queries
• Scalability, efficiency, and reliability
• Sorting results by relevance
• Search in intra- and extra-net
• Improves as GDS does
Issues

- Granularity - files only

- Opaque/proprietary implementation - we do not really know how GDS works
Availability

- Free

- http://ges.sourceforge.net/