

IRiSS – A Source Code Exploration Tool

Denys Poshyvanyk
Andrian Marcus
Yubo Dong
Andrey Sergeyevev

Computer Science Department
Wayne State University
Detroit, MI, USA



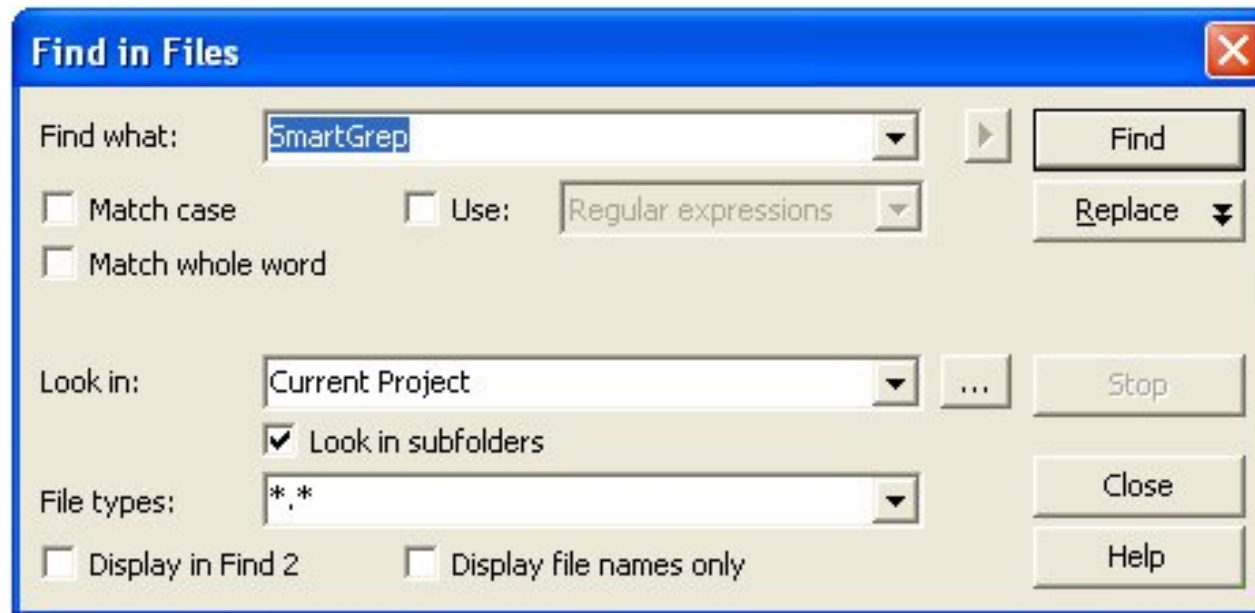
Concept Location

- Locating the implementation of a concept or feature in the source code
- Static
 - Dependency based search [Chen'00]
 - String based search (i.e., grep)
 - IR methods [Marcus'04]
- Dynamic
 - Execution traces - Reconnaissance [Wilde'92]
- Combined (using concept analysis) [Eisenbarth'03]
- Used in incremental change, comprehension, debugging, etc.

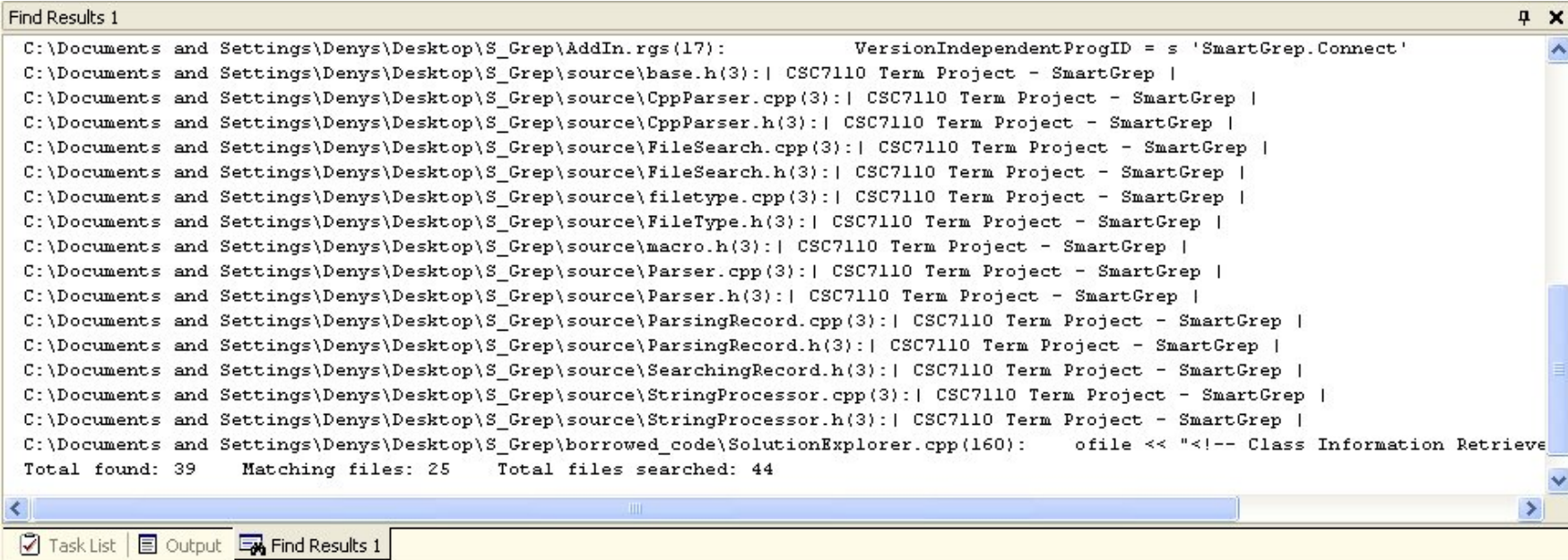
Concept Location

- Locating the implementation of a concept or feature in the source code
- Static
 - Dependency based search [Chen'00]
 - **String based search (i.e., grep)**
 - **IR methods** [Marcus'04]
- Dynamic
 - Execution traces - Reconnaissance [Wilde'92]
- Combined (using concept analysis) [Eisenbarth'03]
- Used in incremental change, comprehension, debugging, etc.

“Find in Files” in .NET?



Results for “Find in Files”



```
Find Results 1
C:\Documents and Settings\Denys\Desktop\S_Grep\AddIn.rgs(17):      VersionIndependentProgID = s 'SmartGrep.Connect'
C:\Documents and Settings\Denys\Desktop\S_Grep\source\base.h(3):| CSC7110 Term Project - SmartGrep |
C:\Documents and Settings\Denys\Desktop\S_Grep\source\CppParser.cpp(3):| CSC7110 Term Project - SmartGrep |
C:\Documents and Settings\Denys\Desktop\S_Grep\source\CppParser.h(3):| CSC7110 Term Project - SmartGrep |
C:\Documents and Settings\Denys\Desktop\S_Grep\source\FileSearch.cpp(3):| CSC7110 Term Project - SmartGrep |
C:\Documents and Settings\Denys\Desktop\S_Grep\source\FileSearch.h(3):| CSC7110 Term Project - SmartGrep |
C:\Documents and Settings\Denys\Desktop\S_Grep\source\filetype.cpp(3):| CSC7110 Term Project - SmartGrep |
C:\Documents and Settings\Denys\Desktop\S_Grep\source\FileType.h(3):| CSC7110 Term Project - SmartGrep |
C:\Documents and Settings\Denys\Desktop\S_Grep\source\macro.h(3):| CSC7110 Term Project - SmartGrep |
C:\Documents and Settings\Denys\Desktop\S_Grep\source\Parser.cpp(3):| CSC7110 Term Project - SmartGrep |
C:\Documents and Settings\Denys\Desktop\S_Grep\source\Parser.h(3):| CSC7110 Term Project - SmartGrep |
C:\Documents and Settings\Denys\Desktop\S_Grep\source\ParsingRecord.cpp(3):| CSC7110 Term Project - SmartGrep |
C:\Documents and Settings\Denys\Desktop\S_Grep\source\ParsingRecord.h(3):| CSC7110 Term Project - SmartGrep |
C:\Documents and Settings\Denys\Desktop\S_Grep\source\SearchingRecord.h(3):| CSC7110 Term Project - SmartGrep |
C:\Documents and Settings\Denys\Desktop\S_Grep\source\StringProcessor.cpp(3):| CSC7110 Term Project - SmartGrep |
C:\Documents and Settings\Denys\Desktop\S_Grep\source\StringProcessor.h(3):| CSC7110 Term Project - SmartGrep |
C:\Documents and Settings\Denys\Desktop\S_Grep\borrowed_code\SolutionExplorer.cpp(160):  ofile << "<!-- Class Information Retrieve
Total found: 39   Matching files: 25   Total files searched: 44
```

Task List | Output | Find Results 1

Our Inspiration

The Google logo is displayed in its signature multi-colored font: 'G' in blue, 'o' in red, 'o' in yellow, 'g' in blue, 'l' in green, and 'e' in red. A small 'TM' trademark symbol is positioned to the upper right of the 'e'.

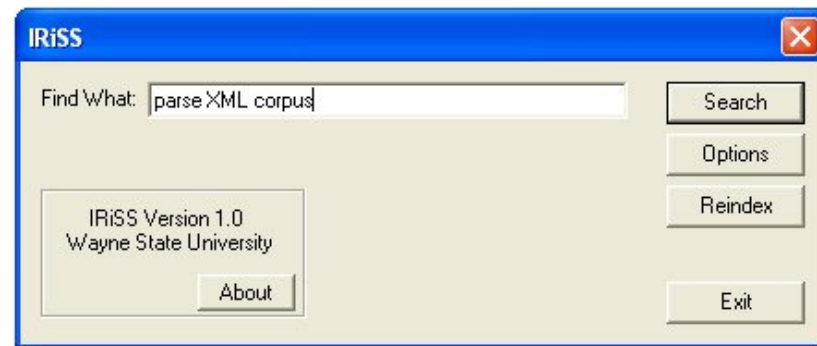
Your Software System

Web [Images](#) [Groups](#) [News](#) [Froogle](#) [Local](#)^{New!} [Desktop](#) [more »](#)

[Advanced Search](#)
[Preferences](#)
[Language Tools](#)

How do We Want to Search?

- “Google-like” queries
- Ranked results
- Various granularity of results (i.e., classes, methods, etc.)



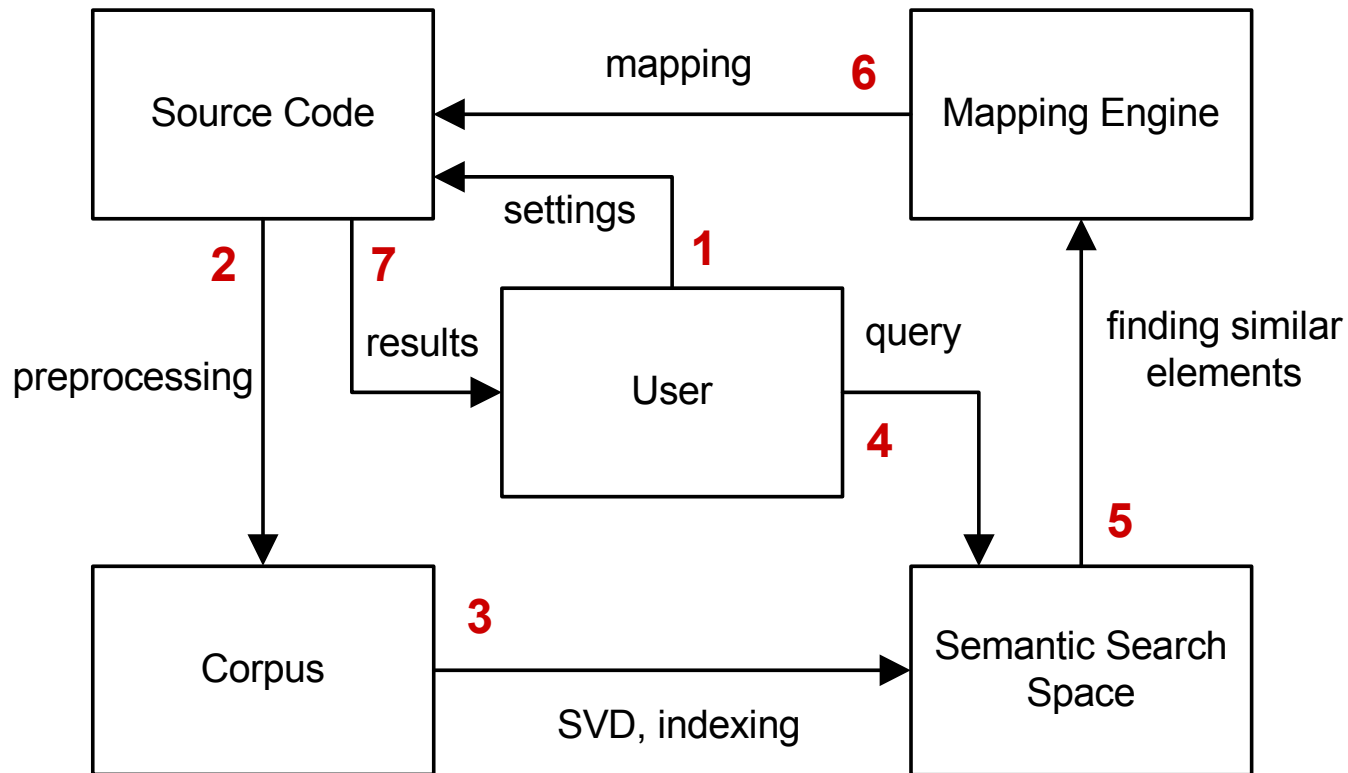
IRiSS

Information **R**etrieval based
Software **S**earching

Information Retrieval

- An Information Retrieval System is capable of storage, retrieval, and maintenance of information (e.g., text, images, audio, video, and other multi-media objects) [Kowalski'97]
- IR methods (e.g., signature files, inversion, clustering, probabilistic classifiers, vector space models, etc.) are used often for text retrieval (e.g., libraries, search engine, etc.)

Searching with IRiSS



Corpus Generation

- Parsing to extract semantic information (i.e., comments and identifiers)
- split_identifiers & SplitIdentifiers
- Define source code documents with user-defined granularity (e.g., class, methods, functions, declarations, interfaces, etc.)
- Works on C/C++
- It is easy to extend to other languages

Building the Semantic Search Space

- We use Latent Semantic Indexing (LSI)
- Each source code element is transformed into a vector, based on the words it contains
- A similarity measure between two documents is defined as the cosine between their corresponding vectors

Query Formulation

- User defined queries
 - Most common, based on user experience and domain knowledge, little known about querying patterns



- Only query terms presented in the corpus are considered

Conclusions

- Pros
 - Simple and flexible to use
 - Returns ranked results – advantage over grep
- Cons
 - Misses some data elements (granularity)
 - Depends on the quality of comments and identifiers (grep has the same problem)
- Using IR methods for concept location and source code browsing is promising

Current Work

- Address the query refinement issue
 - Identify misspelled words
 - Suggest additional words/elements based on first set of results – combine them with AND, OR, XOR, NOT
 - Define the semantic signature of a source code element
 - Deal with scalability issues
- Combine LSI with dependency graph based search, dynamic methods, concept analysis
- Implement the tool as plug-in to Eclipse as well

Google Eclipse Search

Google Desktop Search + Eclipse

The screenshot shows the Eclipse IDE interface with the following components:

- GES Toolbar:** Located at the top left, it contains icons for various Eclipse actions like Save, Run, and Stop.
- Highlighted Source Code:** The main editor window shows the code for `ArrowHeadEditor.java`. The string `"Black Diamond"` in the constructor and the constant `ArrowHead.BLACK DIAMOND` in the `values` array are highlighted in green.
- Search Results:** The Tasks window at the bottom right shows the search results for `"Black diamond"`, indicating 3 matches in the workspace. The results list files in the `Violet.com` package structure.

```
Resource - ArrowHeadEditor.java - Eclipse Platform
File Edit Source Refactor Navigate Search Project Run Window Help
GES Toolbar
Package Explorer
Violet
  com
    horstmann
      violet
        framework
          ActorNode.class
          ActorNode.java
          ArrowHead.class
          ArrowHead.java
          ArrowHeadEditor.class
          ArrowHeadEditor.java
          BentStyle.class
          BentStyle.java
          BentStyleEditor.class
          BentStyleEditor.java
          CallEdge.class
          CallEdge.java
          CallEdgeBeanInfo.class
          CallEdgeBeanInfo.java
          CallNode.class
          CallNode.java
          CallNodeBeanInfo.class
          CallNodeBeanInfo.java
          CircularStateNode.class
          CircularStateNode.java
          ArrowHead.java
          ArrowHeadEditor.java
          ClassDiagramGraph.java

ArrowHeadEditor.java
  "v",
  "Diamond",
  "Black Diamond"
};

private static final Object[] values =
{
  ArrowHead.NONE,
  ArrowHead.TRIANGLE,
  ArrowHead.V,
  ArrowHead.DIAMOND,
  ArrowHead.BLACK DIAMOND
};
};

Tasks Search
"Black diamond" - 3 matches in Workspace
Violet
  com
    horstmann
      violet
        ArrowHead.java
        ArrowHeadEditor.java
        ClassDiagramGraph.java

Writable Smart Insert 21 : 29
```


Future Work – Result Clustering



Vivísimo®

company | products | solutions | customers | demos | press

software engineering the Web [Advanced Search](#)
[Help](#)

NEW search the [Wikipedia](#) at [Clusty.com](#)

Clustered Results Top 225 results of at least **21,270,010** retrieved for the query **software engineering** ([Details](#))

- ▶ [software engineering](#) (225)
 - ▶ [Computer Science](#) (18)
 - ▶ [Software Engineering Research](#) (19)
 - ▶ [Management](#) (16)
 - ▶ [Conference](#) (13)
 - ▶ [Database](#) (13)
 - ▶ [IEEE. Transactions on Software Engineering](#) (10)
 - ▶ [Books](#) (11)
 - ▶ [Solutions](#) (13)
 - ▶ [Resources](#) (12)
 - ▶ [Process](#) (10)
 - ▶ [More](#)

Find in clusters:

1. [Carnegie Mellon - Software Engineering Institute](#) [new window] [frame] [preview] [clusters]
 Read about collaboration opportunities, products and services, publications, and management practices.
[www.sei.cmu.edu](#) - MSN 1, Wisenut 1, Ask Jeeves 1, Open Directory 10, Looksmart 18
2. [TCSE: Technical Council on Software Engineering IEEE-CS](#) [new window] [frame] [preview] [clusters]
 WARNING: Your browser does not support Frames. Please upgrade your browser and try <http://tcse.org> ...
[www.tcse.org](#) - Wisenut 3, Ask Jeeves 11, MSN 12
3. [Software-Engineer.ORG](#) [new window] [frame] [preview] [clusters]
Engineers can discuss training and certification, absorb concepts such as Hierarchical Input Process Output Diagrams and real-time systems, and practice OOP principles.
[www.software-engineer.org](#) - Looksmart 5, MSN 8, Lycos 9, Ask Jeeves 13
4. [Center for Software Engineering](#) [new window] [frame] [preview] [clusters]
 Research at the University of Southern California includes **software** architecture and collaborative **software engineering**.
[sunset.usc.edu](#) - Looksmart 8, Ask Jeeves 14, Open Directory 18, MSN 24
5. [Software Engineering Laboratory - NASA](#) [new window] [frame] [preview] [clusters]
 Partnership developed between NASA, the University of Maryland and Computer Sciences Corporation. Find data, workshops and related sites.
[sel.gsfc.nasa.gov](#) - MSN 3, Ask Jeeves 15, Looksmart 21
6. [RSP&A - Software Engineering Resources](#) [new window] [frame] [preview] [clusters]
 Pressman and Associates offers hundreds of resources related to products and development.
[www.rspa.com/spi](#) - MSN 2, Ask Jeeves 3, Wisenut 23, Looksmart 24

DEMONSTRATION