

Exemplar: <http://www.xemplar.org>

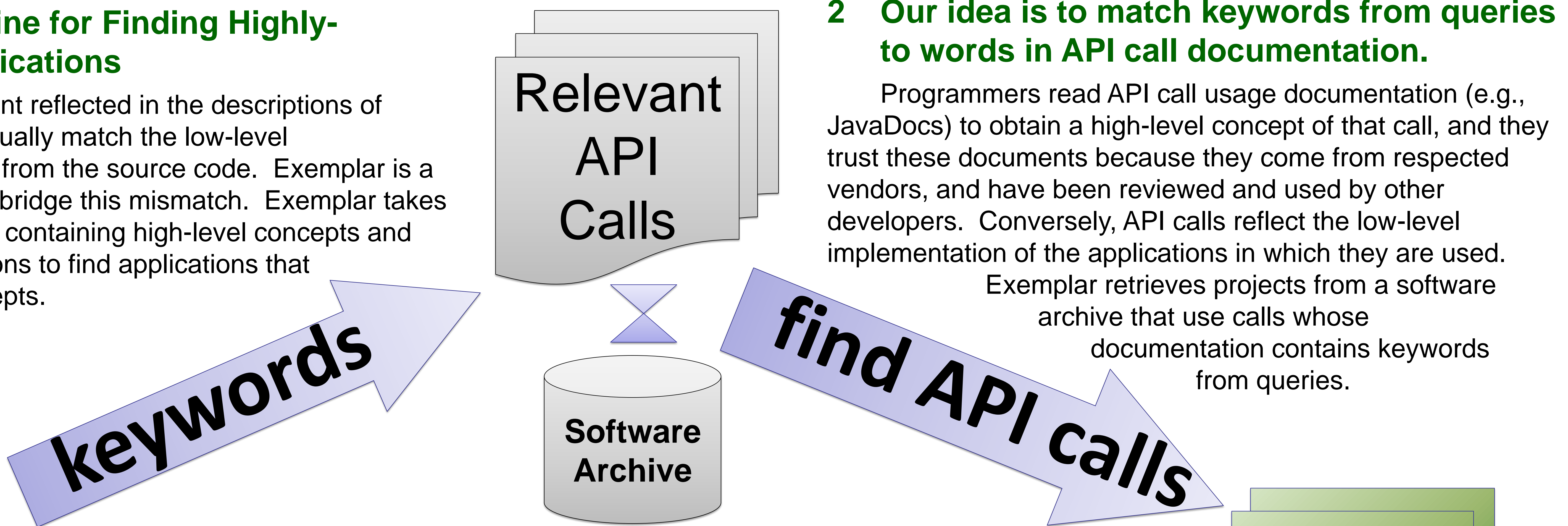
EXEcutable exaMPLeS ARchive

Mark Grechanik^{2,3}, Chen Fu², Qing Xie², Collin McMillan¹, Denys Poshyvanyk¹, Chad Cumby²

¹Computer Science Department, College of William & Mary, ²Accenture Technology Labs, ³Computer Science Department, University of Illinois, Chicago

1 A Search Engine for Finding Highly-Relevant Applications

The high-level intent reflected in the descriptions of applications doesn't usually match the low-level implementation details from the source code. Exemplar is a code search engine to bridge this mismatch. Exemplar takes keywords from queries containing high-level concepts and uses API call descriptions to find applications that implement these concepts.



2 Our idea is to match keywords from queries to words in API call documentation.

Programmers read API call usage documentation (e.g., JavaDocs) to obtain a high-level concept of that call, and they trust these documents because they come from respected vendors, and have been reviewed and used by other developers. Conversely, API calls reflect the low-level implementation of the applications in which they are used.

Exemplar retrieves projects from a software archive that use calls whose documentation contains keywords from queries.

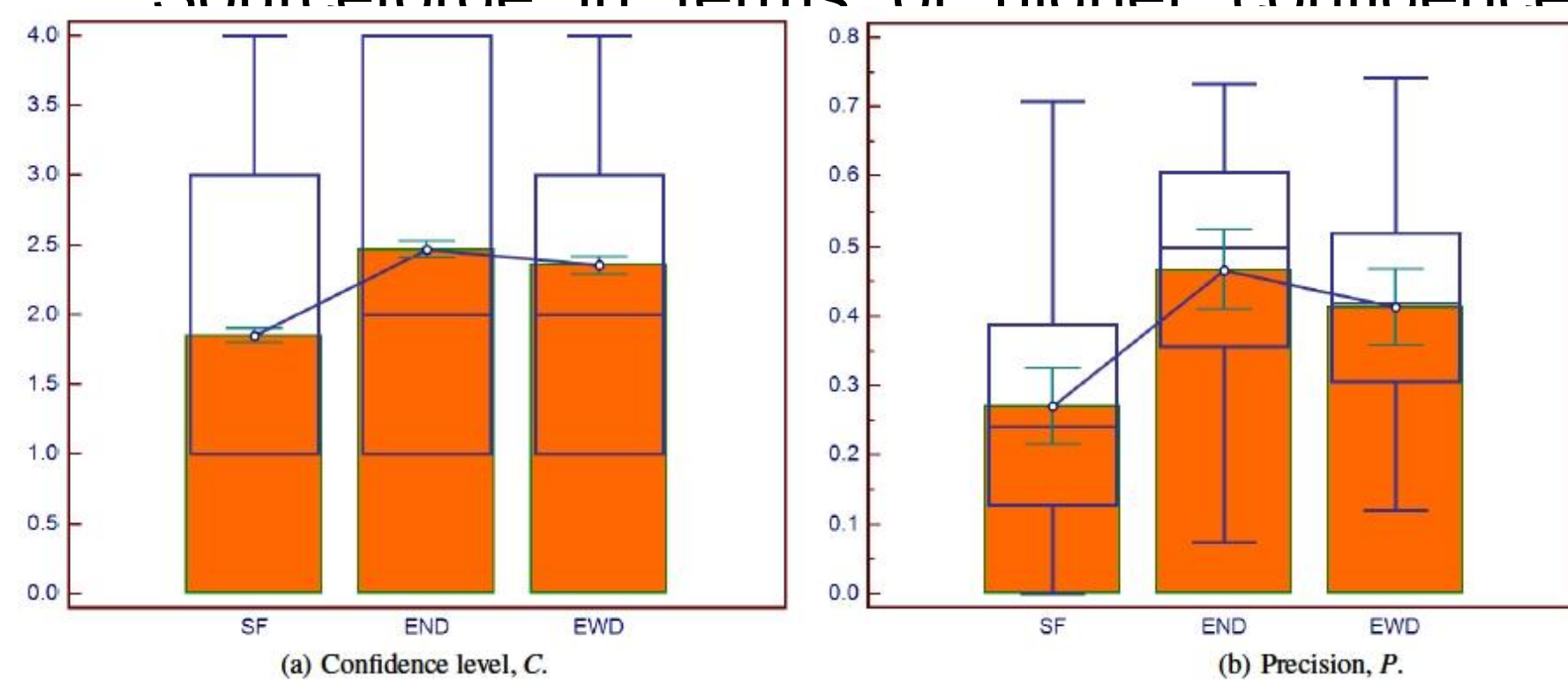
Source Code Crawler

We are building and testing our own source code crawler for downloading, extracting, and indexing open-source applications from repositories, such as Sourceforge.net.

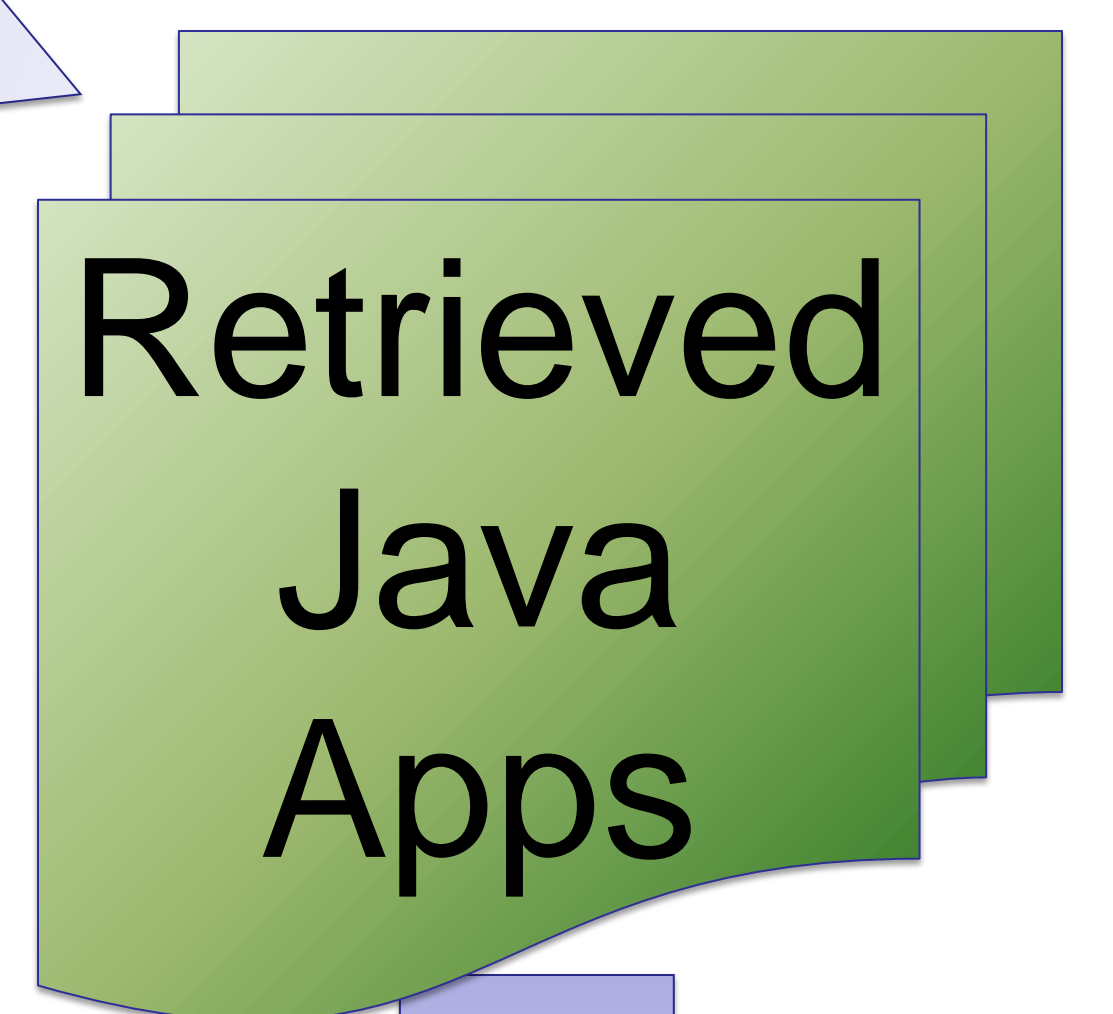
Items	Count
Java Applications	21,934
Files	38,330
Files Downloaded (*.zip, etc.)	31,371
Files Skipped (*.exe, *.pdf, etc.)	6,959
GB Downloaded	105.62 GB
GB Skipped	45.71 GB
Files Indexed by Lucene	10,897
Java docs in index	100,866

Case Study Results

We evaluated Exemplar with 39 professional Java programmers and found with strong statistical significance that it performed better than Sourceforge in terms of higher confidence (C)



SF – Sourceforge
END – Exemplar without dependency information
EWD – Exemplar with dependency information



rank

Ranked Apps List

More ranking

API Call Relevance

Dataflow Connections

Word Occurrences

4 We also determine relations between API calls in retrieved applications.

We observe that relations between concepts entered in queries are often preserved as dataflow links between API calls that implement these concepts in program code. For example, if a user enters keywords **secure** and **send**, and the corresponding API calls **encrypt** and **email** are connected via some dataflow, then an application with these connected API calls is more relevant than one where the calls are not connected.

3 There are three components to compute scores in Exemplar's ranking system.

In addition to API call information, Exemplar ranks applications based on the occurrences of keywords from the query in the high-level descriptions of applications provided by the application's author. Thus, we find applications that match high-level descriptions and low-level implementation details.

Publications

Grechanik, M., Fu, C., Xie, Q., McMillan, C., Poshyvanyk, D., and Cumby, C., "Exemplar: EXEcutable exaMPLeS ARchive", in Proceedings of 32nd ACM/IEEE International Conference on Software Engineering (ICSE'10), Formal Research Tool Demonstration, Cape Town, South Africa, May 2-8, 2010.

Grechanik, M., Fu, C., Xie, Q., McMillan, C., Poshyvanyk, D., and Cumby, C., "A Search Engine For Finding Highly Relevant Applications", in Proceedings of 32nd ACM/IEEE International Conference on Software Engineering (ICSE'10), Cape Town, South Africa, May 2-8, 2010.

Resources

Exemplar is publicly available! <http://www.xemplar.org/>

