

Understanding Code Search Intents from Web Search Engine

Jin-woo Park, Sanghoon Lee, Seung-won Hwang

Speaker : **Jin-woo Park**

Information and Database Systems Laboratory
Department of Computer Science & Engineering
Pohang University of Science and Technology (POSTECH)
jwpark85@postech.edu

Outline

1. Introduction
2. Preliminaries
3. Analysis
4. Approach
5. Q & A

Outline

1. Introduction

2. Preliminaries

3. Analysis

4. Approach

5. Q & A

Motivation



He wants to find the JAVA information from general search engine?

Motivation



query

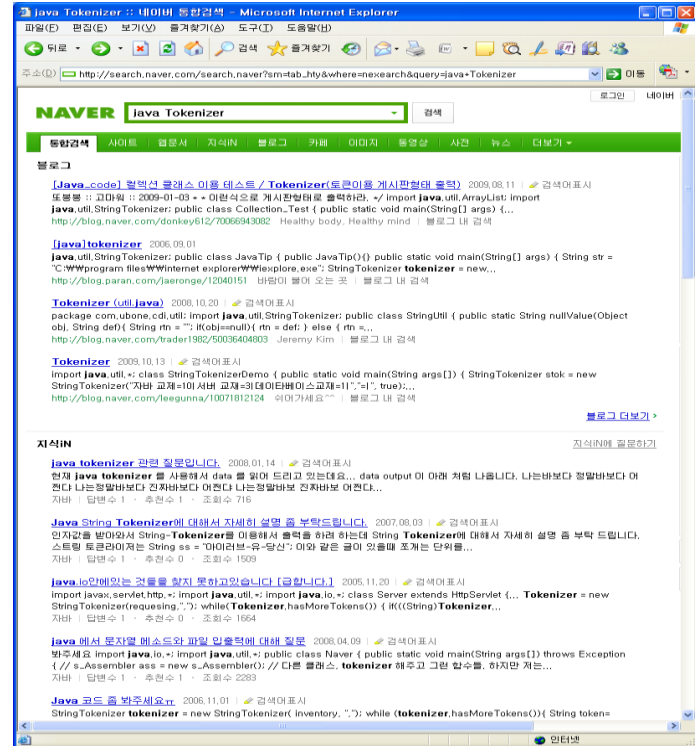


He wants to find the JAVA information from general search engine?

Motivation



query

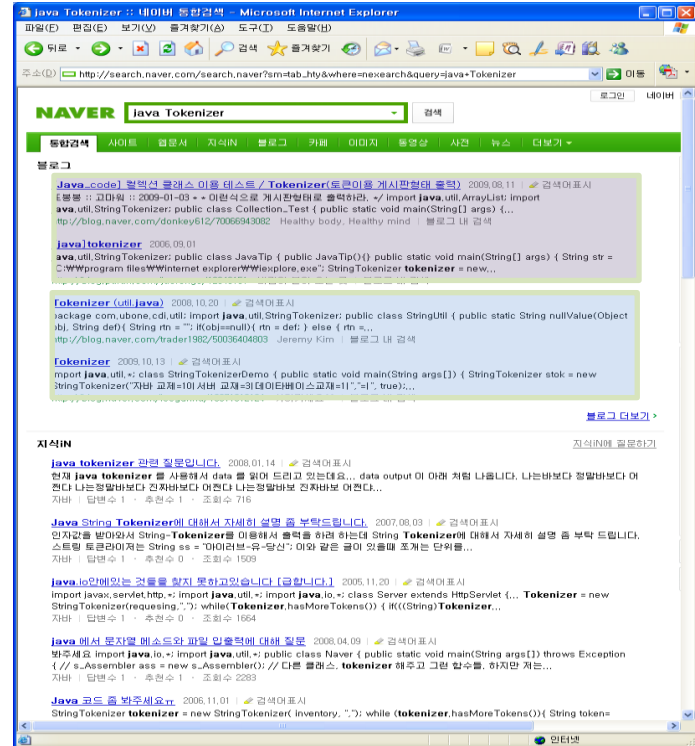


He wants to find the JAVA information from general search engine?

Motivation



query



He wants to find the JAVA information from general search engine?

Motivation



query

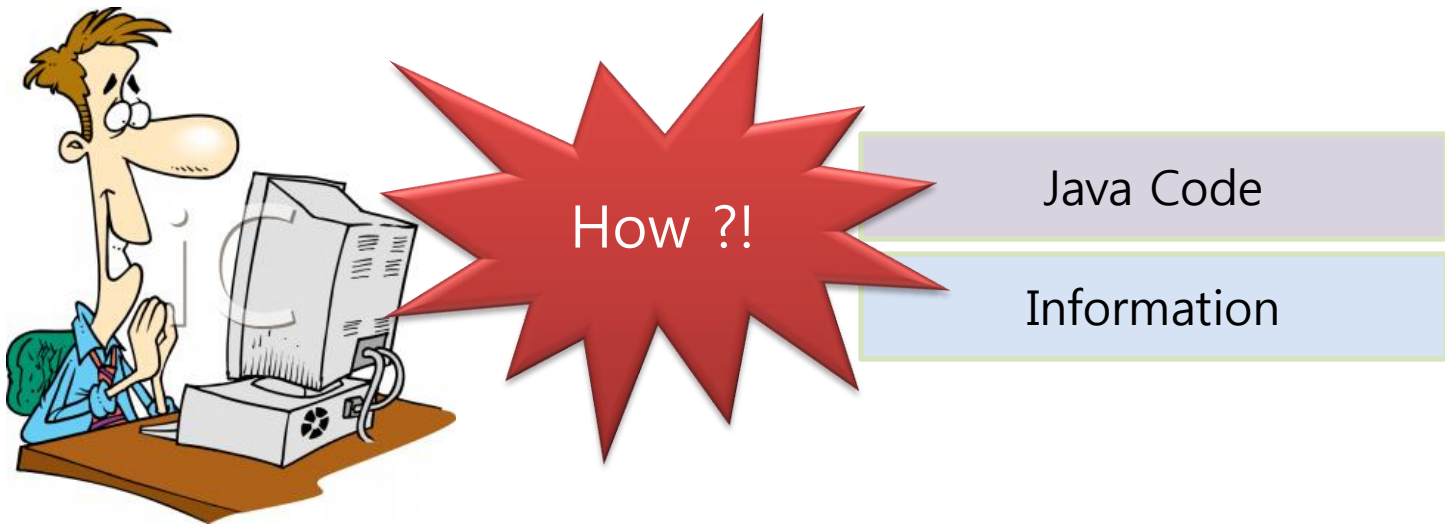


Java Code

Information

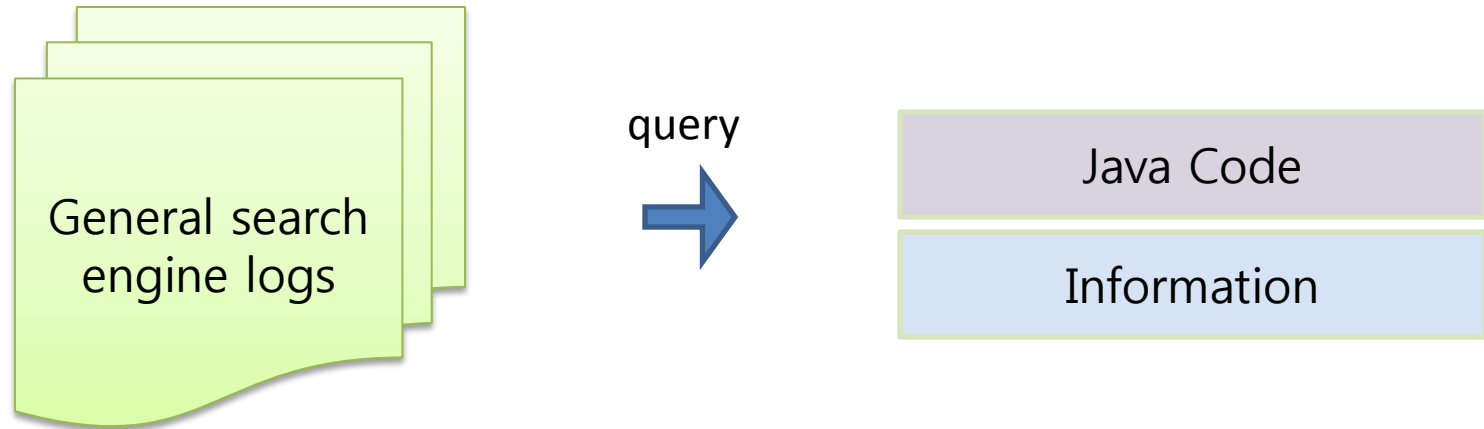
What is his search intent?

Motivation



What is his search intent?

Overview



What is his search intent?

Goal

Java Code



Java Tokenizer :: 네이버 동원검색 - Microsoft Internet Explorer

주소(D) http://search.naver.com/search.naver?sm=tab_hhty&where=nexearch&query=java+Tokenizer

NAVER Java Tokenizer 검색

동원검색 사이트 방문시 지시인 블로그 카페 이미지 동영상 사진 뉴스 더보기

별문시

- String (Java Platform SE 6)**
additional information on **string** concatenation and conversion, see Gosling, Joy, and Steele, The **Java Language Specification**. Unless otherwise noted, passing a **null** argument to a...
<http://java.sun.com/javase/6/docs/api/java/lang/String.html> 사이트 내 검색 | 저장된 페이지
- String (Java 2 Platform SE 5.0)**
additional information on **string** concatenation and conversion, see Gosling, Joy, and Steele, The **Java Language Specification**. Unless otherwise noted, passing a **null** argument to a...
<http://java.sun.com/j2se/1.5.0/docs/api/java/lang/String.html> 사이트 내 검색 | 저장된 페이지
- Java Byte Array To String**
orbok though, is simply, french hospital san luis obispo "Is it possible to create a **java string** from a byte array if that byte array. Private static javalangstring...
<http://download.potak.xorg.pl/salless.html> 사이트 내 검색 | 저장된 페이지
- Scala for Java Refugees Part 11: main(String[]) - Code Commit**
Array[String]. That is to say, args is a **string** array. In Scala, **Array** is actually a **class** (a real class, not like **Java** arrays) that takes a type parameter defining the...
<http://www.codecommit.com/blog/scala/scala-for-java-refugees-...> 사이트 내 검색 | 저장된 페이지
- String (Java 2 Platform SE v1.4.0)**
append(**Java.lang.String**). CharSet. Serialized FormField Summary static Comparator CASE_INSENSITIVE_ORDER A Comparator...
<http://java.kecc.co.kr/j2se/1.4/docs/api/java/lang/String.html> 사이트 내 검색 | 저장된 페이지

원문시 더보기

블로그

- Max length of a String in JAVA?** 2006.11.01 | 검색어포시
So I guess that's the max length of a **String**. Edwin Marlin
-----**Java** character takes two bytes, so correct...
<http://blog.naver.com/care01hk/150010432311> vaki's home 블로그 내 검색
- ...value of type 'Java.lang.String' to be applied to a List or array** 2006.11.24 | 검색어포시
The ":" operator was supplied with an index value of type **Java.lang.String** to be applied to a List or array... JSTL을 사용하면서 이와 같은 예외가 발생할 경우 다음의 링크를 참조하자. ...
<http://blog.naver.com/korcstewis/110011283310> Revolution 2007 | 블로그 내 검색
- [Java] String 클래스** 2009.12.31 | 검색어포시
reportViewTest(ReportTestCid.java:77) - 같아요~~~~ 문자열의 시작... **String** 객체 생성 **String.copyValueOf(data, offset, count)** Returns a **String** that represents the character sequence in the array specified. data ...
<http://blog.naver.com/minis24/80098511196> 제너럴의 열거말 스토리 | 블로그 내 검색
- [Java] String format 그리고 한글** 2009.12.19

안료 인터넷

Goal

Information

A screenshot of a NAVER search results page for the query "Java Tokenizer". The page shows several search results with titles, dates, and snippets of code. The search bar at the top contains "Java Tokenizer" and the search button is labeled "검색". The results are listed under the heading "블로그" (Blog).

[Java_code] 원형식 클래스 스터용 레소트 / Tokenizer(호른이름 게시판형의 동원) 2008.08.11 | 검색어포시
도움말 | 고대원 | 2009-01-05 = - 이원식으용 게시판형태를 출력하러. </ import java.util.ArrayList; import java.util.StringTokenizer; public class Collection_Test { public static void main(String[] args) { ... } http://blog.naver.com/donkey612/70066943082 Healthy body, Healthy mind | 블로그 내 검색

[Java]tokenizer 2006.09.01
Java.util.StringTokenizer; public class JavaTip { public JavaTip(){} public static void main(String[] args) { String str = "C:\\Program files\\WInternet explorer\\Wwexplore.exe"; StringTokenizer tokenizer = new... http://blog.paran.com/jaeronge/12040151 | 바람이 불어 오는 곳 | 블로그 내 검색

Tokenizer (util.java) 2008.10.20 | 검색어포시
package com.ubone.cd.util; import java.util.StringTokenizer; public class StringUtil { public static String nullValue(Object obj, String def) { String str = ""; if(obj==null) { str = def; } else { str =... http://blog.naver.com/trader1982/5003640303 Jeremy Kim | 블로그 내 검색

Tokenizer 2009.10.19 | 검색어포시
import java.util.*; class StringTokenizerDemo { public static void main(String args[]) { StringTokenizer stok = new StringTokenizer("자바 교재=이서바 교재=이데터바=이교재=이"); boolean b = true;... http://blog.naver.com/leegunna/10071812124 | 이이개세요 | 블로그 내 검색

자바 | 답변수 1 | 추천수 0 | 조회수 1684

Java 토큰라이저는 String ss = "이이러브=우=당신"; 이와 같은 글에 있을때 포개는 단위를... 2007.08.09 | 검색어포시
인자값을 받아와서 String-Tokenizer를 이용해서 출력할 하러 하러던 String-Tokenizer에 대해서 자세히 설명 좀 부탁 드립니다. 스트링 토큰라이저는 String ss = "이이러브=우=당신"; 이와 같은 글에 있을때 포개는 단위를... http://blog.naver.com/leegunna/10071812124 | 이이개세요 | 블로그 내 검색

Java 로인애있는 것들을 찾지 못하고있습니다.(급한니다) 2005.11.20 | 검색어포시
import javax.servlet.http.*; import java.util.*; import java.io.*; class Server extends HttpServlet { ... Tokenizer = new StringTokenizer(requesting...); while(Tokenizer.hasMoreTokens()) { IK((String)Tokenizer... http://blog.naver.com/leegunna/10071812124 | 이이개세요 | 블로그 내 검색

Java 에서 문자열 메소드와 파일 입출력에 대해 질문 2008.04.09 | 검색어포시
방주세요 import java.io.*; import java.util.*; public class Naver { public static void main(String args[]) throws Exception { // s.Assembler ass = new s.Assembler(); // 다른 클래스. tokenizer 해주고 그런 함수를, 하지만 저는... http://blog.naver.com/leegunna/10071812124 | 이이개세요 | 블로그 내 검색

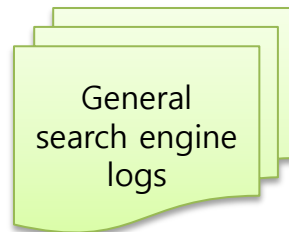
Java 코드 좀 봐주세요 2006.11.01 | 검색어포시
StringTokenizer tokenizer = new StringTokenizer("inventor..."); while (tokenizer.hasMoreTokens()) { String tokens

Goal

Java Code



Explanation



Outline

1. Introduction

2. Preliminaries

3. Analysis

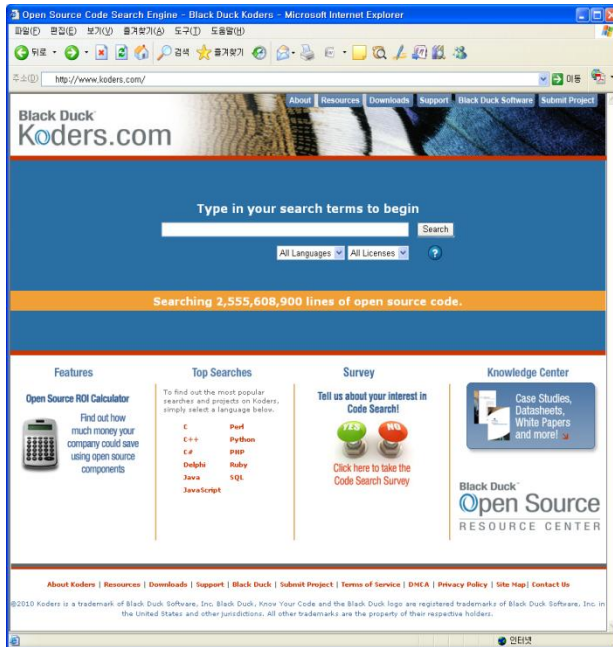
4. Approach

5. Q & A

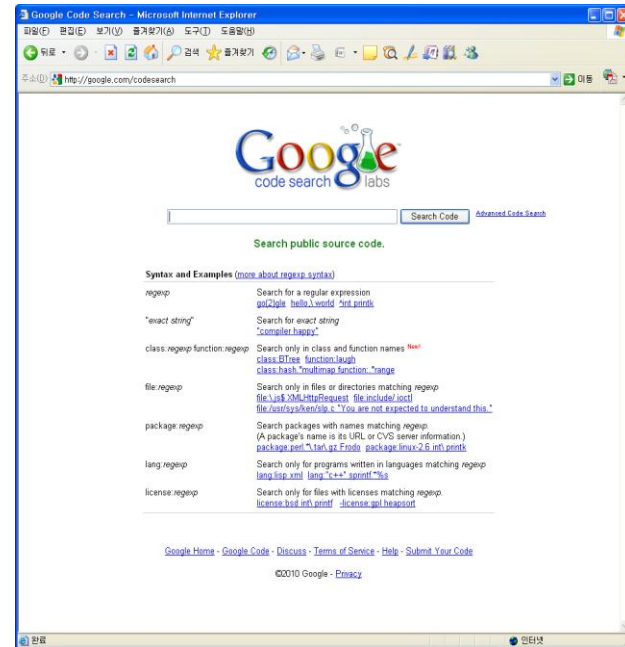
Preliminaries

- Code search engines

 - Koders



 - Google code search



Preliminaries

- Code search engines

- 1. Advantages**

- Exemplifying the actual usage of the given method
 - Expediting software development process

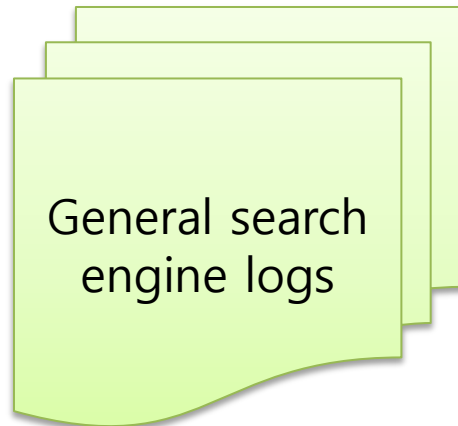
- 2. Problems**

- Difficult to express a query
 - Searching only source codes

Preliminaries

- Log from web search engine
 - ▶ **Log of User-click behavior**

<query ID, query, URL, position of results page, time-stamp>



Preliminaries

- Mining search categories

Topic Code	Description	Words	Topic Code	Description	Words
apache	Terms from Apache projects	org, apache, commons, ant, axis, stringutils, poi, lang, tools, impl, beanutils, catalina,	junit	Junit	test, exception, junit, framework, queue, testcase, font, start, assert, tapestry, antlr,
audio	Working with audio and sound	compare, control, encode, audio, decode, uuid, diff, sound, record, barcode, rectangle,	log4j	Logging with log4j	log, logger, log4j, configuration, logging, config, common, logfactory, info, debug, util,
BIRT	BIRT project from Eclipse	birt, ibm, matrix, page, report, group, scanner, int, ws, a, width, decimalformat,	lucene	Lucene, Information Retrieval	lucene, main, hash, javascript, index, field, count, function, displaytag, mozilla, analysis,
calSched	Calendar and scheduling	calendar, password, excel, quartz, sorting, codehaus, xfire, none, gregoriancalendar,	M.GisThr	Mixed topics: Unknown	thread, ldap, geotools, manager, wrapper, cvs, monitor, point, draw, polygon, lock, jndi,
DataStr	Data structures	list, object, arraylist, map, vector, hashmap, array, set, add, iterator, hashtable, collection,	M.GwtSecAuth	Mixed topics: Possibly login/authentication	google, login, user, gwt, example, authentication, agent, grid, checkbox,
DateTime	Date and Time	tree, date, time, format, btree, binary, simpledateformat, b, timestamp, avi,	M.rng	Random numbers and timing	random, number, calculator, zip, timer, generator, port, edu, serial, clock, packet, cs,
db.DAO	Data Access Objects, Object Relational Mapping	table, name, dao, base, getproperty, stringtokenizer, column, persistence, ibatis,	M.util	Mixed topics: possibly utilities	util, new, properties, pattern, load, enum, regex, bufferedreader, constants, stringutil,
db.SJMO	Database	sql, jdbc, connection, database, mysql, oracle, datasource, update, tag, select, pool,	media	Multimedia	type, error, video, media, player, mp3, template, serializable, change, classloader,
DEF	Queries with operators, examples in Koders	cdef, mdef, idef, insert, tree, mcall, compute, account, getuserid, jme, ruby, controller, dir,	ML	WEKA, Machine Learning	m, math, huffman, weka, attribute, instance, call, language, linear, fibonacci, x, entry,
Docs	Various document formats	text, pdf, dom, document, print, node, element, xpath, lowagie, w3c, jdom, itext,	Netbeans	Netbeans	org, netbeans, editor, api, command, modules, windows, lib, openide, console,
Eclipse	Eclipse project	eclipse, org, ui,.swt, internal, core, jface, jdt, resources, widgets, wizard, osgi, gef,	Network	Networking, FTP	client, server, ftp, socket, iso, 3166, smtp, chat, tcp, jxta, udp, ftpclient, nio,
files	Working with files	file, read, files, create, write, upload, csv, directory, copy, input, reader, open, save,	NL.apps	Natural language queries for projects/apps	system, project, application, management, koders, c, sip, h, projects, library, out, online,
fqn.J3DSfCon	Mixed FQN fragments from sun, sourceforge,	com, net, sun, util, sf, utils, tools, j3d, sourceforge, internal, concurrent, bea, edu,	NL.HUD	Natural language, "How to" type queries	and, using, from, data, the, get, how, with, program, value, find, for, display, check, not
gameMob	Mobile games and GUI	game, j2me, bluetooth, mobile, sms, microedition, midlet, chess, canvas, polish,	NL.SC	Natural language queries for "source code"	code, source, for, mail, algorithm, network, email, send, sample, puyo, internet, codes,
GUI	Swing and AWT GUI	swing, jtable, applet, awt, window, menu, event, jtree, snmp, frame, button, dialog,	parsing	Parsers and Compilers	parser, parse, base64, compiler, parsing, htmlparser, funambol, xmlparser, utility,
hibernate	Hibernate project	hibernate, session, address, property, query, transaction, id, mapping, lookup, expression,	searchEng	Searching	search, engine, rmi, browser, method, first, status, remote, reflect, desktop, invoke,
http	Http and URLs	http, download, url, proxy, httpclient, protocol, line, httpURLConnection, rss, ping,	secAuth	Security and Authentication	security, jar, service, jasper, ssl, auth, engine, jasperreports, export, dori, signature,
imaging	Image processing, image file formats	image, color, jpeg, compression, tiff, processing, jai, bufferedimage, fft, paint,	sort	Sorting, Koder's example terms	sort, md5, card, merge, quicksort, shell, messenger, multi, poker, hello, selection, join,
JAVA	java.lang.io, using 'java' in query terms	java, lang, io, koders, jav, noclasseffounderror, doctor, soapdient,	Spring	Spring Framework	org, web, springframework, spring, context, factory, support, ajax, view, beans, services,
jboss	Jboss project	org, jboss, ejb, cache, resource, loader, j2ee, spi, sudoku, jmx, container, weblogic,	streams	Working with streams	default, inputstream, process, size, content, runtime, length, buffer, response,
jfreeC	JFreeChart charting library	html, jfreechart, chart, jfree, dataset, data, wicket, talnet, link, renderer, demo, dynamic,	String	Working with Strings	string, convert, 1, byte, int, array, replace, 0, split, converter, 2, character, stringbuffer,
Jkw	Java Keywords	class, extends, public, key, implements, gnu, crypto, cipher, des, rsa, interface, void,	Struts	Struts project	struts, filter, action, form, validator, validation, validate, opensymphony, bean, actionform,
JMS	Java Message Service	message, jms, listener, toString, integer, resourcebundle, mcall, locale, xsd, datatype,	U1	Random words, unknown Topic	import, word, bouncycastle, header, ioexception, provider, velocity, utils, location,
JsfJgraph	JSF and Jgraph	javax, graph, faces, component, model, jsf, wsdl, event, portlet, jgraph, myfaces, help,	U2	Unknown, possibly FQN fragments	package, com, gui, plugin, task, all, v2, xstream, basic, mchange, impl, c3p0,
JspServ	JSP and Servlets	servlet, jsp, request, path, javax, httpServletRequest, tomcat, multipart,	XmLWS	XML and web services	xml, shopping, cart, javax, soap, rpc, sax, bind, transform, webservice, schema, stream,

Preliminaries

- Mining search categories

Category	Topics under category
Applications	calSched, gameMob, media, searchEng
Programming Tasks	audio, DataStr, DateTime, db.DAO, Docs, files, http, M.rng, Network, parsing, secAuth, sort, streams, String, XmlWS
Form Centric Java/JDK Libraries	DEF, fqj.j3DSfCon, Jkw, NL.apps, NL.HUD, NL.SC db.SJMO, GUI, imaging, JAVA, JMS
Frameworks	apache, BIRT, Eclipse, hibernate, jboss, jfreeC, JsfJgraph, JspServ, junit, log4j, lucene, ML, Netbeans, Spring, Struts
Unknown	M.GisThr, M.GwtSecAuth, M.util, U1, U2

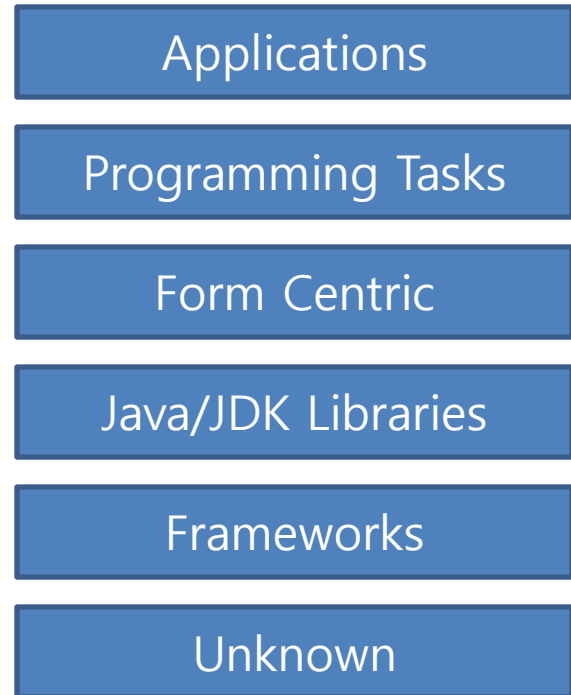
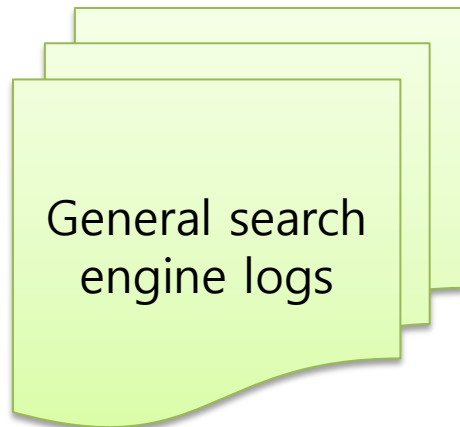
- **Koders queries on Java were classified into categories**

Outline

1. Introduction
2. Preliminaries
- 3. Analysis**
4. Approach
5. Q & A

Analysis

- Categorizing logs based on the categories



<for loop java programming, <http://java.sun.com/docs/books/tutorial/java/nutsandbolts/for.html>>

< free online java games, <http://www.arcademachine.com/>>

⋮

<using java image effects, <http://www.frontlook.com/jfx/jfxhelp/toc.htm>>

Analysis

- Categorizing logs based on the categories

Category	Koders	General
Applications	7.1%	34.9%
Programming Tasks	29.9%	20.2%
Form Centric	13.5%	20.2%
Java/JDK Libraries	12.7%	5.4%
Frameworks	28.4%	15.4%

Topic frequency comparison of each category

Analysis

- Understanding search intents

Applications

Programming Tasks

Form Centric

Java/JDK Libraries

Frameworks

Java parser



The screenshot shows a web page titled "The for Statement (The Java™ Tutorial: Learning the Java Language) | Language Basics | Microsoft Internet Explorer". The page content includes:

Notice how the code declares a variable within the initialization expression. The scope of this variable extends from its declaration to the end of the block governed by the for statement, so it can be used in the termination and increment expressions as well. If the variable that controls a for statement is not needed outside of the loop, it's best to declare the variable in the initialization expression. The names `i`, `j`, and `k` are often used to control for loops; declaring them within the initialization-expression limits their life span and reduces errors.

The three expressions of the for loop are optional: an infinite loop can be created as follows:

```
for ( ; ; ) { // infinite loop
    // your code goes here
}
```

The for statement also has another form designed for iteration through [Collections](#) and [arrays](#). This form is sometimes referred to as the *enhanced* for statement, and can be used to make your loops more compact and easy to read. To demonstrate, consider the following array, which holds the numbers 1 through 10:

```
int[] numbers = {1,2,3,4,5,6,7,8,9,10};
```

The following program, [EnhancedForDemo](#), uses the enhanced for to loop through the array:

```
class EnhancedForDemo {
    public static void main(String[] args){
        int[] numbers = {1,2,3,4,5,6,7,8,9,10};
        for (int item : numbers) {
            System.out.println("Count is: " + item);
        }
    }
}
```

In this example, the variable `item` holds the current value from the numbers array. The output from this program is the same as below:

```
Count is: 1
Count is: 2
Count is: 3
Count is: 4
Count is: 5
Count is: 6
Count is: 7
Count is: 8
Count is: 9
Count is: 10
```

We recommend using this form of the for statement instead of the general form whenever possible.

Analysis

- Understanding search intents

Category	Navigational queries(%)
Programming Tasks	53.57%
Frameworks	17.86%
Applications	3.57%
Overall	8%

Navigational query ratio of dominant categories

Outline

1. Introduction
2. Preliminaries
3. Analysis
- 4. Approach**
5. Q & A

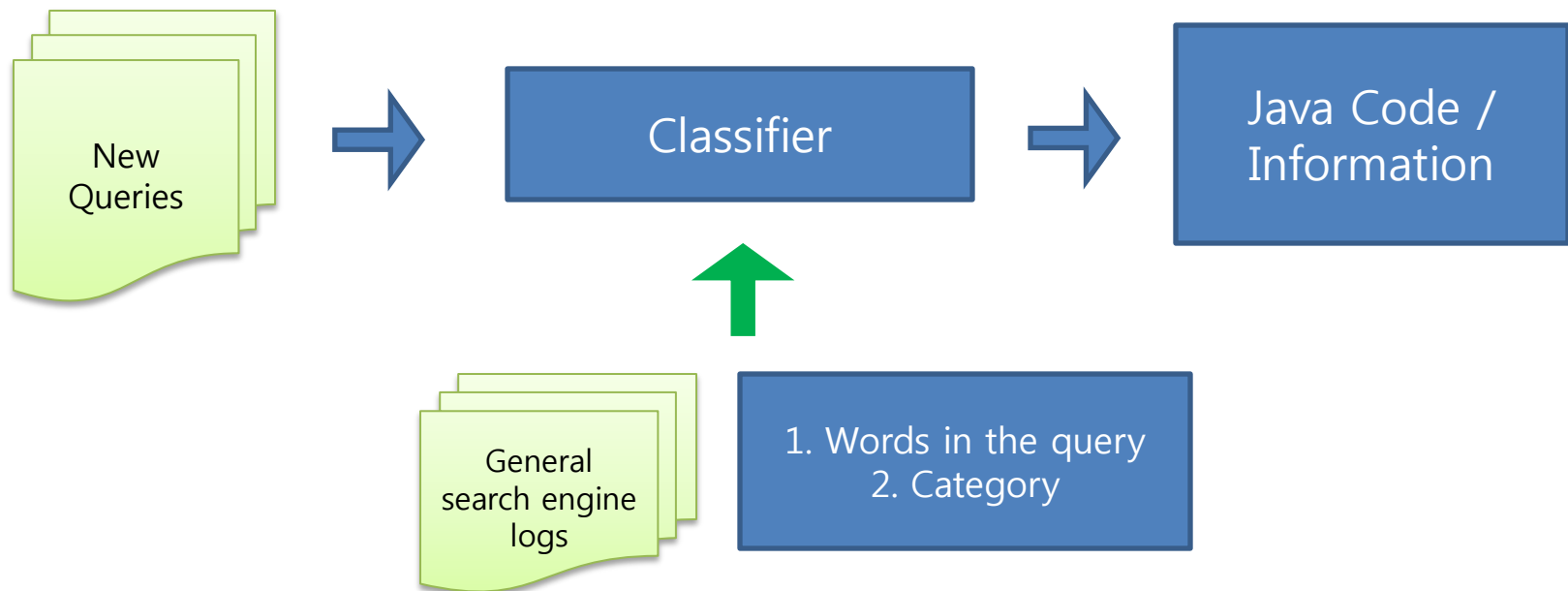
Approach

- Search Intents Prediction



Approach

- Search Intents Prediction

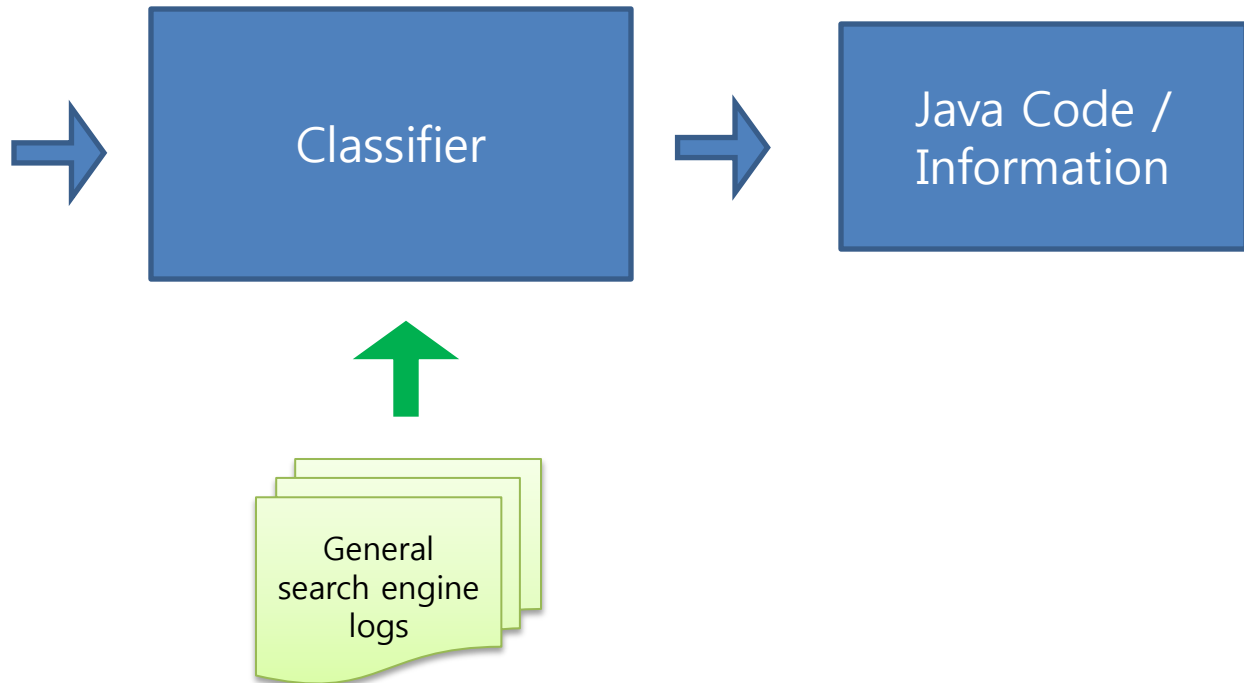


Conclusion

Java Code



Explanation



Q & A

Thank you!