REALSEARCH

ABB
Power and productivity
for a better world™

TMAP: Discovering Likely Mapping between APIs using Text Mining

Rahul Pandita*, Raoul Jetley**, Sithu Sudarsan**, and Laurie Williams*

*North Carolina State University, USA
**ABB Corporate Research, India

Motivation

- Developers need to release different versions of their software
    - Each platform has its own APIs, classes, methods – Functionality similar across these
    - Migration from one platform to another
    - Need to identify best match for API calls between platforms

- Existing methods
    - Manual translation & mapping
    - Statistical language translation (StaMiner)
    - Method Alignment (MAM)
    - Dynamic approach (Rosetta)
    - Information retrieval (web search) (Zheng et. al.)

TMAP: Discovering Likely API Mappings using Text Mining
Research Question

How accurately can the similarity in the language of API method descriptions be leveraged to discover likely API Mappings?

---

drawString

```
public void drawString(String str,
                       int x,
                       int y,
                       int anchor)
```

Draws the specified `string` using the current font and color. The `x,y` position is the position of the anchor point. See anchor points.

Parameters:
    `str` - the `String` to be drawn
    `x` - the x coordinate of the anchor point
    `y` - the y coordinate of the anchor point
    `anchor` - the anchor point for positioning the text
Throws:
    `NullPointerException` - if `str` is null
    `IllegalArgumentException` - if anchor is not a legal value
See Also:
    `drawChars(char[], int, int, int, int, int)`

---

public void **drawText** (String text, float x, float y, Paint paint)

Draw the text, with origin at (x,y), using the specified paint. The origin is interpreted based on the Align setting in the paint.

Parameters

text   The text to be drawn

x   The x-coordinate of the origin of the text being drawn

y   The y-coordinate of the baseline of the text being drawn

paint   The paint used for the text (e.g. color, size, style)

TMAP: Discovering Likely API Mappings using Text Mining
Research Question

- Simply searching the API documentation for a relevant map not practical
  - Using words in method signature, got >23,500 results in Android API

- Problems
  - Presence of keywords – public, void, static, etc.
  - Generic method names – toString(), get(), set()
  - Not all terms have equal weights
  - Structure of API documents

TMAP: Discovering Likely API Mappings using Text Mining
Approach
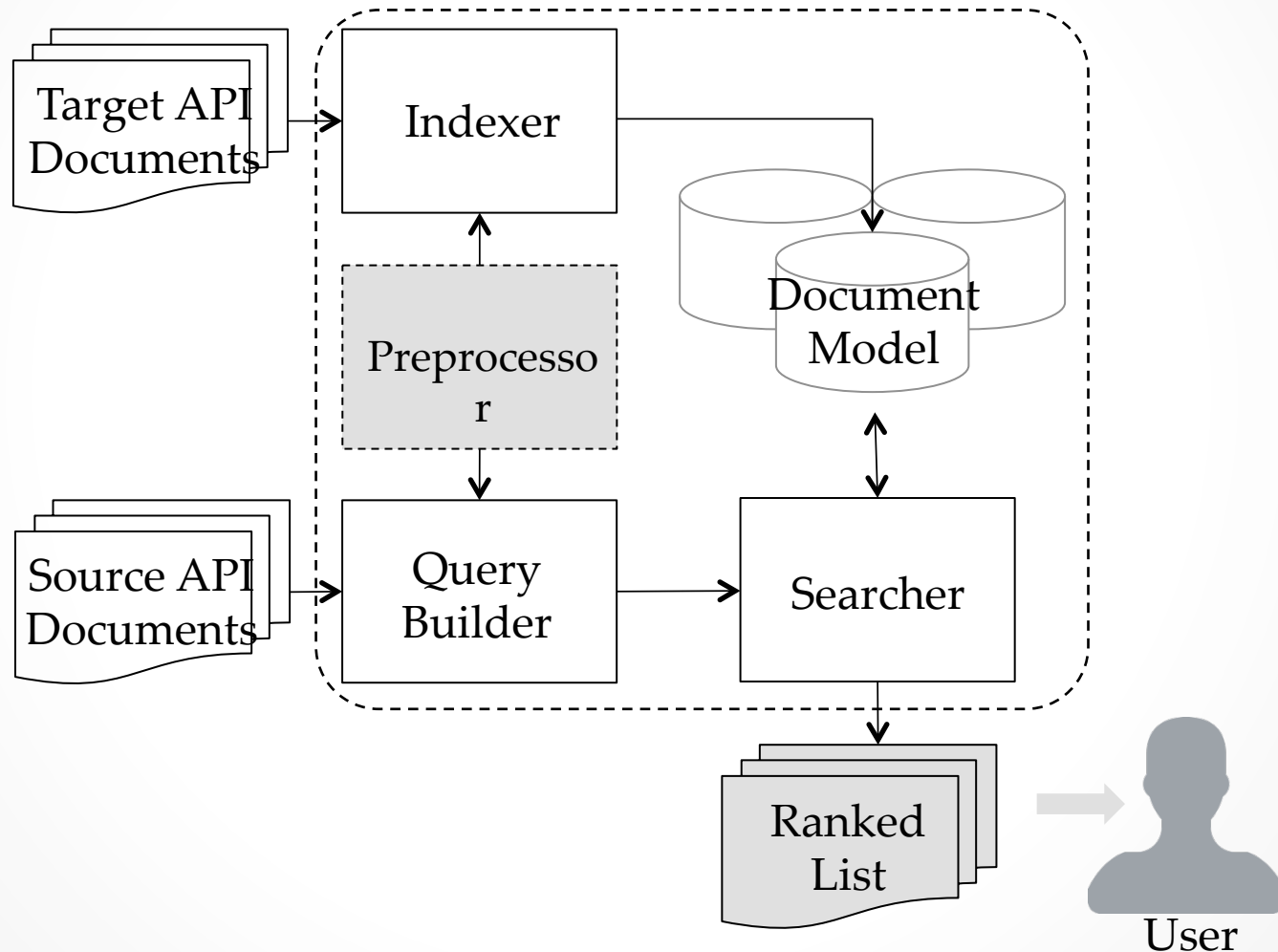
TMAP: Discovering Likely API Mappings using Text Mining
Approach

- Indexer - accepts the API documents of the target API & and creates indexes for these documents

    o Extracts intermediate contents from the method descriptions

- Query Builder – accepts the API documents of the source API and creates queries to be executed on the indexes

- Searcher – executes queries on indexes and generates a ranked list of API mappings from target API documents

TMAP: Discovering Likely API Mappings using Text Mining
Indexer

- Index contains
    - Type name
    - Package name
    - Method name
    - Class description
    - Method description

- Removal of stopwords, annotations (@Code, @link)
- Splitting of packages: javax.microedition.lcdui → javax microedition lcdui
- Lower case translation
- Splitting of CamelCase notation: drawstring → draw string
- Stemming: Porter stemmer

TMAP: Discovering Likely API Mappings using Text Mining
Query Builder

- Builds queries to search index
    - Derives information from source API to search target API index
    - Search on type definitions, method names

- Heuristic 1: *the first paragraph or the first five sentences of the type description (whichever is shorter) provides reasonable keywords for searching equivalent class in target API*
- Heuristic 2: *the first paragraph or the first two sentences of the method description (whichever is shorter) provides reasonable keywords for searching equivalent method in target API*

- Example: hasNext → "ha next"

TMAP: Discovering Likely API Mappings using Text Mining
Searcher

- Use Lucene API for Searching
- Use Cosine similarity to rank results

- Cosine similarity
  - Numerical statistic to measure the similarity between two vectors. Cosine similarity is defined as a dot product of magnitude of two vectors.
  - In text mining, Cosine similarity is used to capture similarity between two documents represented as term frequency vectors

TMAP: Discovering Likely API Mappings using Text Mining
Evaluation

- Used to get mappings for
  - JavaME to Android (5 classes)
  - Java to C# (10 classes)

- Mappings categorized as
  - Relevant
  - Exact

- coverage (Cov) – ratio of the number of methods in a type that TMAP found *at-least* one relevant mapping to the total number of source methods in that type.
  - $\triangle$Cov – Increase in Cov in comparison to results reported by previous approaches (Rosetta and StaMiner),
- $\triangle$Cov = TMAP$_{cov}$–Prev$_{cov}$.
  - High value of $\triangle$Cov indicates the effectiveness of TMAP in finding API method mappings

TMAP: Discovering Likely API Mappings using Text Mining

| S No. | API | | Type | No. Mtds | Relevant | | Exact | | ΔCov | Common | New |
	Source	Target			Prev	TMAP	Prev	TMAP			
1	JavaME	Android	javax.microedition.lcdui.Alert	16	3	15	3	7	0.75	0	7
2	JavaME	Android	javax.microedition.lcdui.Canvas	22	5	18	5	10	0.6	0	10
3	JavaME	Android	javax.microedition.lcdui.Command	6	3	3	3	0	0	0	0
4	JavaME	Android	javax.microedition.lcdui.Graphics	39	18	36	18	29	0.47	5	24
5	JavaME	Android	javax.microedition.lcdui.Font	16	3	15	3	8	0.75	0	8
6	Java	C#	java.io.File	54	15	37	15	26	0.41	7	19
7	Java	C#	java.io.Reader	10	1	8	1	6	0.7	1	5
8	Java	C#	java.io.Writer	10	2	10	2	10	0.8	1	9
9	Java	C#	java.util.Calendar	47	0	11	0	5	0.24	0	5
10	Java	C#	java.util.Iterator	3	0	3	0	1	1	0	1
11	Java	C#	java.util.HashMap	17	5	9	5	5	0.24	1	4
12	Java	C#	java.util.ArrayList	28	6	22	6	15	0.58	4	11
13	Java	C#	java.sql.Connection	52	1	28	1	13	0.52	1	12
14	Java	C#	java.sql.ResultSet	187	10	146	10	31	0.73	1	30
15	Java	C#	java.sql.Statement	42	1	21	1	5	0.48	1	4
Total				549	73	382	73	171	0.57	22	149

TMAP: Discovering Likely API Mappings using Text Mining