

MAPPING THE RESPONSES OF RESTFUL SERVICES BASED ON THEIR VALUES

Blerina Bazelli, Marios Fokaefs, Eleni Stroulia
University of Alberta, Canada

(Not always) A match made in heaven

- A client-service relationship may break when:
 - A web service becomes unavailable
 - It becomes insufficient for the clients' needs
 - It changes frequently over time
 - Its price increases
- The client will migrate to another service.

A Wealth of Similar(?) APIs



7 years

6,000 APIs

3,000 news stories

Three steps for API migration

- Discovery
 - ▣ Comparing the desired API against available ones
 - ▣ To recognize if the differences are possible to adapt
- Selection
 - ▣ Comparing the desired API against a set of discovered ones
 - ▣ To identify the “best” one to use
- Mapping
 - ▣ Comparing two similar APIs
 - ▣ To precisely decide how to adapt one to match the other (for composition or for substitution)

How to Recognize API Substitutability?

Previous Approaches

Based on the similarity of

- signature syntax
 - ▣ Same operations, with same number and types of input/output data
- signature specification
 - ▣ Same pre- and post-conditions
- signature semantics
 - ▣ Synonymous identifiers (operations, parameters)

How to Recognize API Substitutability?

Our Approach

Based on the similarity of

- the input and output values of their corresponding invocations
- If two parameters have the same instance value, they refer to the same thing.

Why responses?

- REST → resource-based services
- Request Data → Analysis → Functionality

The Methodology

Interactive Steps →see Future Directions

1. Identify two substitutable APIs
2. Map their corresponding operations
3. Invoke the corresponding operations pairwise, with the same parameter values

Automated Step

4. Use the proposed algorithm to precisely map the output parameters

Comparison Rules –numbers

$$\text{dist}(n_1, n_2) = \frac{\max(|n_1|, |n_2|) - \min(|n_1|, |n_2|)}{\max(|n_1|, |n_2|)}, \text{ (eq. 1)}$$

where $|n_1|, |n_2|$ are the absolute values of two numbers n_1, n_2

□ $n_1 = 106$

□ $n_2 = 93$

distance = 100 (13/106) = 12.2%

Comparison Rules –strings

$$\begin{aligned} & lev_{a,b}(i, j) && , \min(i, j) = 0 \\ = & \begin{cases} \max(i, j) & , \text{ else} \\ \min \begin{cases} lev_{a,b}(i - 1, j) + 1 \\ lev_{a,b}(i, j - 1) + 1 \\ lev_{a,b}(i - 1, j - 1) + [a_i \neq b_j] \end{cases} \end{cases} \end{aligned}$$

kitten → sitting

- **k**itten → **s**itten (substitution of "s" for "k")
- **s**it**e**n → **s**it**i**n (substitution of "i" for "e")
- **s**it**i**n → **s**it**i**ng (insertion of "g" at the end)

distance = 100(3/7) = 42.8%

Comparison Rules –bag of words

$$\text{similarity}(S1, S2) = 100\left(1 - \frac{S1 \cap S2}{\min(|S1|, |S2|)}\right)$$

where $S1, S2$ two sets containing string data types and $|S1|, |S2|$ the number of elements in $S1$ and $S2$ respectively.

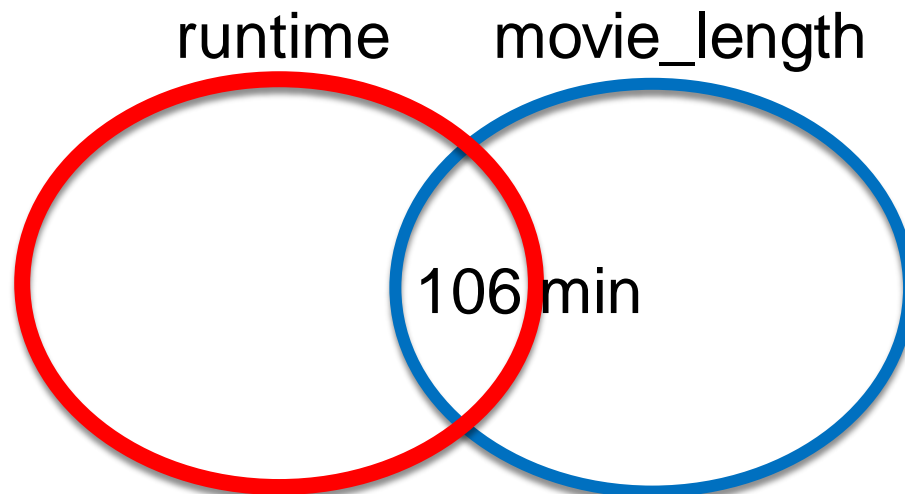
□ $S1 = \{a,b,d,e\}$

□ $S2 = \{a,b,f,g,h\}$

distance = $100(1-2/4) = 50\%$

Comparison Rules –inclusion

- When comparing two elements,
- If one is a prefix or a suffix of the other,
- Then they are similar



Heuristics

- String vs. String
1. Transform to lowercase
 2. Compute Levenshtein distance

String vs. Number □ Extract the number sequences

Number vs. Number □ Maximum computed similarity

Date vs. Date □ Inclusion

Evaluation

- When is this method applicable?
- When the APIs refer to a **commonly shared domain of factual information**
 - ▣ Non factual → judgment may cause differentiation to the values of entities
 - ▣ Proprietary → the values of entities will not be common across APIs

Evaluation

Domains

1. Geolocation

Web Services

- Google Maps
- Microsoft Maps
- CloudMade Maps

2. Movies

- Internet Movie Database (IMDb)
- Rotten Tomatoes
- Filmaster
- TheMovieDB

Results

Type	Source Element	Source Value	Target Element	Target Value	Similarity (%)
Set-Set	name	[Carrie Underwood, Lorraine Nicholson, Annasophia Robb, Helen Hunt, Dennis Quaid]	Actors	[Carrie Underwood, Chris Brochu, AnnaSophia Robb, Sonya Balmores, Lorraine Nicholson, Craig T. Nelson, Dennis Quaid, Helen Hunt...]	100.00
Date-Date	release_date	20110408	Theater	2011-04-08	100.00
String-Integer	runtime	USA: 106 min	Runtime	106	100.00
Integer-Integer	release_year	2011	Year	2011	100.00
String-String	title_localized	Soul Surfer	Title	Soul surfer	100.00
Double-Double	average_score	5.95	Rating	6.9	86.23
Integer-String	imdb	1596346	imdb_id	tt1596346	100.00
Integer-String	id	771037147	imdb_id	tt1596346	20.00
String-String	surname	McNamara	Type	M	12.50
String-String	length	null	runtime	USA: 106min	0.00
Double-Double	coordinates	[-113.5177994, -113.5173264, 53.5225983]	lng	-113.518802	99.98

First-best mapping: Precision & Recall

$$\textit{Precision} = \frac{TP}{TP + FP}$$

$$\textit{Recall} = \frac{TP}{TP + FN}$$

Source & Target elements that have been mapped **correctly**

Mapped only manually

Elements wrongly mapped as correct

Results – Geolocation

Source \ Target	Google					Bing					CloudMade				
	TP	FP	FN	P (%)	R(%)	TP	FP	FN	P(%)	R(%)	TP	FP	FN	P(%)	R(%)
Google						10	1	1	90.91	90.91	2	3	1	40.00	66.67
Bing	13	3	1	81.25	92.86						11	2	0	84.62	100.00
CloudMade	2	4	1	33.33	66.67	9	3	1	75.00	90.00					

Results – Movies

Source \ Target		IMDb					RottenTomatoes				
		TP	FP	FN	P (%)	R(%)	TP	FP	FN	P(%)	R(%)
IMDb							7	3	3	70.00	70.70
Rotten Tomatoes		8	0	2	100.00	80.00					
Filmaster		11	4	3	73.33	78.57	4	5	3	44.44	57.14
TheMovieDB		5	2	1	71.43	83.33	3	6	1	33.33	75.00
Source \ Target		Filmaster					TheMovieDB				
		TP	FP	FN	P (%)	R(%)	TP	FP	FN	P(%)	R(%)
IMDb		9	2	2	81.82	81.82	3	6	0	33.33	100.00
Rotten Tomatoes		4	2	3	66.67	57.14	3	3	1	50.00	75.00
Filmaster							5	3	1	62.5	83.33
TheMovieDB		4	5	2	44.44	66.67					

Many-to-many mapping :Mean Average Precision

$$MAP = \frac{\sum_{p=1}^P AvePre(p)}{P}, eq(2)$$

where p is a pair of elements and P the number of elements that can be mapped

- p : pair of elements
- P : number of elements that can be mapped
- AvePre: average precision per element

MAP: Geolocation

Source \ Target	Google	Bing	CloudMade
	P (%)	P(%)	P(%)
Google		79.52	25.00
Bing	58.62		17.85
CloudMade	20.00	30.83	

MAP: Movies

Source \ Target	IMDB	RottenTomatoes	Filmaster	TheMovieDB
	P (%)	P(%)	P(%)	P(%)
IMDB		57.16	67.00	67.00
Rotten Tomatoes	58.33		38.00	35.00
Filmaster	75.00	57.14		80.00
TheMovieDB	33.33	50.00	74.00	

Future Directions

- Combine more methods of comparison
 - ▣ Instance value mapping
 - ▣ Structure mapping
 - ▣ Lexicographic and semantic mapping
- Automate as many steps of the process as possible.
 - ▣ Use a service discovery and selection method.
 - ▣ Extend the mapping process to input values to map the operations.
 - ▣ Automatically exercise the APIs (testing?)

Conclusions

- A methodology for mapping services based on their run-time invocations
- For Web services with factual and shared information, the evaluation demonstrated high precision & recall
- The objective is to simplify the effort required to develop the adapters to migrate from one service to its substitute