Synonym Discovery for Bing Local

ZANWAR FARAJ

AUGUST 20, 2013

MANAGER: MURALI BRAHMADESAM

BING DOMAINS – ONLINE SERVICES DIVISION A.K.A. APPLICATIONS & SERVICES

Background

Business Synonym: Alternative name for the same or a closely related business

Types of Synonyms

- *Strong Synonyms:* Refer to the same business
    - Ex: Subway ↔ Subway Inc
- *Partial Synonyms:* Refer to a closely related business
    - Ex: Sears ↔ Sears Auto Center

Windows 8.1 Search Charm

- Now connected to Bing
- Search for local businesses with Bing without opening web browser

Search

Share

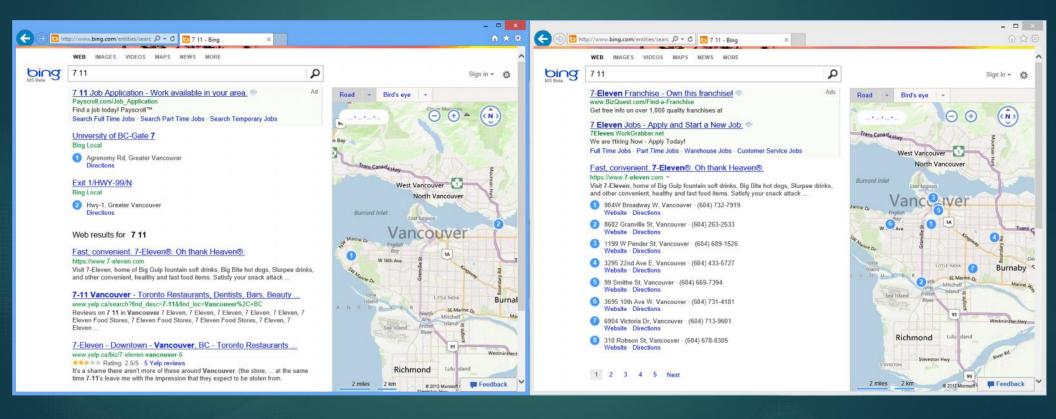Start

Devices

Settings

Objective

Improve Bing Local search results through synonym discovery

- Add business synonyms to Bing's database
  - Provides more information for the search engine to operate

- Produce results for previously ignored queries
  - Example: Show results if user queries "7 11" instead of "7-Eleven"

Before

After

Additional Applications

Business Name Normalization

▶ Determine a standard name for synonymous entities

Canadian Tire, Div of
Canadian Tire Corp
Canadian Tire Corporation
Canadian Tire Ltd
Canadian Tire Store

→ | Canadian Tire |

▶ Improve user experience

  ▶ Local results should show the most popular name for a particular business

Additional Applications (cont.)

Website Composition

▶ Use synonym database to propagate website information between synonymous entities

Alberta Wood Products	http://www.albertawood.com/
Alberta Wood	No Website

→

Alberta Wood Products	http://www.albertawood.com/
Alberta Wood	http://www.albertawood.com/

▶ Provide more information about queried entities

   ▶ Can be applied to other fields (ex: phone numbers)

   ▶ Directly increases Bing's LDCG

   *LDCG is a metric that indicates the quality of search results

Synonym Discovery Techniques

Strong Synonyms

Acronyms/Expansions

▶ Liquor Control Board of Ontario ↔ LCBO

▶ ACA ↔ Alberta Conservation Association

Unigrams (differs by one word)

▶ Tim Horton's ↔ Tim Horton's <u>Café</u>

Bigrams (differs by two words)

▶ Total Image ↔ Total Image <u>Beauty Salon</u>

Partial Synonyms

Partial Acronyms/Expansions

▶ <u>CCA</u> ↔ <u>Canada Christian Academy</u> Brampton Inc

▶ <u>Canadian Standards Association</u> ↔ <u>CSA</u> Group

Partial Matches

▶ <u>Esso</u> ↔ Virginia-Park <u>Esso</u>

Numerical Prefixes/Suffixes

▶ L.C.B.O. ↔ L.C.B.O. <u>#126</u>

▶ Subset of partial matches (indicates store #)

Results

Synonym Discovery

	Type	Biz Names	Synonyms	Precision
Strong Synonyms	Acronyms/Expansions	1,341	1,600	88%
	Unigrams	109,411	119,074	93%
	Bigrams	4,651	4,768	91%
	Total	114,001	125,437	90%
Partial Synonyms	Partial Acronyms/Expansions	2,232	3,527	86%
	Partial Matches	62,869	107,304	100%
	Numerical Prefixes/Suffixes	2,084	4,292	100%
	Total	64,306	110,350	93%
All Synonyms	Total	178,307	235,787	

*Totals are not a simple sum of each individual count as duplicate synonyms must be removed

Name Normalization

Strong Synonyms

- ▶ Distinct Business Names Normalized: *20,099*
- ▶ Distinct Entities Normalized: *38,910*

Partial synonyms not used in name normalization as a normalized name should only be applied if the synonymous entities are interchangeable

Website Composition

Strong Synonyms

- ▶ Distinct Entities Affected: *19,672*

Partial Synonyms

- ▶ Distinct Entities Affected: *29,473*

Conclusion

▶ Summary of Synonym Discovery

  ▶ Improve search results

  ▶ Normalize business names

  ▶ Share website information between entities

▶ Future discovery techniques

  ▶ Translation (using Bling API)

  ▶ Cosine similarity (numerical indicator of the similarity between two business names)

  *Bling is a translation API service offered by Microsoft