Discovering Subsumption Relationships for Web-Based Ontologies

Dana Movshovitz-Attias, Steven Whang, Natasha Noy, Alon Halevy
7:00 am – 5:00 pm
Brother Baba Budan, Saturday hours
Brother Baba Budan - Cafes - Melbourne Victoria - Yelp
www.yelp.com › Restaurants › Cafes › Yelp
★★★★★ Rating: 4.5 - 71 reviews - Price range: $
71 Reviews of Brother Baba Budan "Breakfast at this great coffee shop. ... Get a seat if you can, because the space is really cool, and try a pastry while you're at ..."

Brother Baba Budan - CBD - Melbourne | Urbanspoon
www.urbanspoon.com › Melbourne › City › CBD › Urbanspoon
★★★★★ Rating: 91% - 2,494 votes
... Blog Posts about Brother Baba Budan, CBD. Brother Baba Budan reviews on Urbanspoon. ... My Legendary Girlfriend. Interesting Venue Names - Melbourne.

Brother Baba Budan - Broadsheet
Brother Baba Budan, named after the 17th-century Sufi who smuggled seven seeds of coffee out of ... The space is small, with a single communal table, a tiny bar and a total of about 15 seats (not including the ones on the ceiling); Budan is, ...

Brother Baba Budan, Melbourne - TripAdvisor
www.tripadvisor.com.au › ... › Melbourne Restaurants › TripAdvisor LLC
★★★★★ Rating: 4.5 - 88 reviews
Brother Baba Budan, Melbourne: See 88 unbiased reviews of Brother Baba Budan, rated 4.5 of 5 on TripAdvisor and ... Very busy take out business with about six seats around a table. ... Nice coffee served in a tiny space on Little Bourke.

Brother Baba Budan, Melbourne - Beanhunter
www.beanhunter.com/australia/victoria/melbourne/.../brother-baba-budan
★★★★★ Rating: 4.3 - 362 reviews
Brother Baba Budan, Best coffee shops & cafes in Melbourne CBD, This is my ... Very busy little place, probably need a bit more seating as was very cramped, ... Missing: eaepeity

Brother Baba Budan (@brotherbudan) | Twitter
https://twitter.com/brotherbudan
The latest Tweets from Brother Baba Budan (@brotherbudan). ... @brotherbudan
Coffee out of ... The space is small, with a single communal table, a tiny bar and a total of about 15 seats (not including the ones on the ceiling); Budan is, ...
Is-A Relations

- restaurant
  - coffee shop
    - brother baba budan
  - fast food restaurant
    - mcdonalds
Is-A Relations

- coffee shop
  - hours
- restaurant
  - seating capacity
- fast food restaurant
  - seating capacity
- brother baba budan
- mcdonalds
Subsumption Relations

- coffee shop
- brother baba budan

- fast food restaurant
- mcdonalds

- restaurant
  - hours
  - seating capacity
Subsumption Relations

Brother Baba Budan

Coffee shop

Restaurant

Fast food restaurant

Seating capacity

Hours

Seating capacity

McDonalds
Strict Subsumption

restaurant

coffee shop

Formal Definition:
\[ C_1 \sqsubseteq C_2 \iff \forall c \in C_1, c \in C_2 \]
Every instance of the sub-concept also belongs to the super-concept
Strict Subsumption

Formal Definition:
\[ C_1 \subseteq C_2 \text{ iff } \forall c \in C_1, c \in C_2 \]

Every instance of the sub-concept also belongs to the super-concept.

Wikipedia:
Coffee shops, while similar to cafés, are not restaurants due to the fact that they primarily serve and derive the majority of their revenue from hot drinks.

Merriam-Webster:
Coffee shop: a small restaurant that serves coffee and other drinks as well as simple foods.
## Strict Subsumption

Users search coffee shop as if it were a restaurant

| restaurant | Formal Definition: $C1 \sqsubseteq C2 \text{ iff } \forall c \in C1, c \in C2$  
|            | Every instance of the sub-concept also belongs to the super-concept |
| coffee shop | Wikipedia: Coffee shops, while similar to cafés, are not restaurants due to the fact that they primarily serve and derive the majority of their revenue from hot drinks. |
|            | Merriam-Webster: Coffee shop: a small restaurant that serves coffee and other drinks as well as simple foods |
Latent Subsumption

Users search coffee shop as if it were a restaurant

Formal Definition:
\[ C_1 \subseteq C_2 \text{ iff } \forall c \in C_1, c \in C_2 \]
Every instance of the sub-concept also belongs to the super-concept

Wikipedia:
Coffee shops, while similar to cafés, are not restaurants due to the fact that they primarily serve and derive the majority of their revenue from hot drinks.

Merriam-Webster:
Coffee shop: a small restaurant that serves coffee and other drinks as well as simple foods
Baseline 1: Dependency Parsing

restaurant

fast food restaurant
Baseline 1: Dependency Parsing

restaurant

fast food restaurant
Baseline 1: Dependency Parsing

restaurant

fast food restaurant

fast
JJ
ADJ

food
NN
NOUN
Number=sing

restaurant
NN
NOUN
Number=sing
Baseline 1: Dependency Parsing

restaurant

coffee shop
Baseline 1: Dependency Parsing

restaurant

coffee shop

**coffee**

- **nn**
- **NN**
- **NOUN**
- **Number=sing**

**shop**

- **nn**
- **ROOT**
- **NN**
- **NOUN**
- **Number=sing**
Baseline 2: Word Vectors

restaurant

coffee shop
Baseline 2: Word Vectors

restaurant

A restaurant called Little Tokyo ...

I'll meet you at the restaurant at 7:15 ...

you've never been to a restaurant before ...

coffee shop

A coffee shop called Jury on the old Pentridge ...

I'll meet you at the coffee shop and ...

You've never been to a coffee shop like this one ...
Baseline 2: Word Vectors

**restaurant**

A *restaurant* called Little Tokyo ...

I'll meet you at the *restaurant* at 7:15 ...

you've never been to a *restaurant* before ...

**word vector 1**

**coffee shop**

A *coffee shop* called Jury on the old Pentridge ...

I'll meet you at the *coffee shop* and ...

You've never been to a *coffee shop* like this one ...

**word vector 2**
Baseline 2: Word Vectors

A **restaurant** called Little Tokyo ...

*I'll meet you at the **restaurant** at 7:15 ...*

you've never been to a **restaurant** before ...

---

A **coffee shop** called Jury on the old Pentridge ...

*I'll meet you at the **coffee shop** and ...*

You've never been to a **coffee shop** like this one ...

---

(Compare)
Using Attributes

restaurant

coffee shop
Using Attributes

restaurant

coffee shop

seating capacity
waitress
chef
advertisement
soft drink
annual income
location
opening times
nutritional information
food menu
coffee
roaster
brew methods
music
jobs

advertisement

opening times

nutritional information
Using Attributes

restaurant

(coffee shop)

(very likely)

seating capacity

waitress

chef

advertisement

soft drink

annual income

location

opening times

nutritional information

coffee menu

music

coffe

roaster

brew methods
Biperpedia [VLDB 14]

- Mines thousands of attributes for every concept from search queries and Web text
  - > 2 orders of magnitude larger than Freebase
  - Attributes ranked around 100K for country: surveillance authorities, aerospace experts, antitrust chief, human resources firm, airport owner, ...

- Also identifies synonyms and mis-spells
  - tourist attraction = tourist spot
  - moter: mother or motor (depending on Person or Car)
<table>
<thead>
<tr>
<th>fast food restaurant</th>
<th>restaurant</th>
<th>coffee shop</th>
<th>school</th>
</tr>
</thead>
<tbody>
<tr>
<td>Latte</td>
<td></td>
<td></td>
<td></td>
</tr>
</tbody>
</table>


Latte

fast food restaurant  restaurant  coffee shop  school

Generate Candidates (using overlapping attributes)

restaurant

fast food restaurant  restaurant  school

coffee shop  restaurant
Generate Candidates (using overlapping attributes)

Generate Features (Attributes, Dependency Parse, Word Vectors) and classify candidates
Attribute Features

● Basic Features
  ○ How “similar” are the concepts?
  ○ How “specialized” is one concept to another?

● Variants
  ○ Different weights for attributes
    ■ frequency, instances, rank
  ○ Synonym-based attribute matching
Experimental Setting

● Sources of concepts
  ○ Is-A Network (47K)
    ■ Apply patterns (e.g., ‘X is a Y’, ‘Y such as X’) on the Web
    ■ Discard nodes with fewer than 10 children
  ○ Freebase (10K)

● Training on Is-A Network concepts
  ○ 5.6K manually-labeled subsumption relations
  ○ Agreement among (three) experts was low: 74%
Precision vs. Recall for different methods:

- **DepParse (F1=0.63)**
- **WordVectors (F1=0.66)**
- **WordVectors + DepParse (F1=0.67)**
- **Attributes (F1=0.71)**
- **All (F1=0.74)**
Large-scale Prediction

2.4M predicted relations

<table>
<thead>
<tr>
<th>Top-k Samples:</th>
<th>100</th>
<th>1K</th>
<th>10K</th>
<th>100K</th>
</tr>
</thead>
<tbody>
<tr>
<td>Estimated Precision:</td>
<td>1</td>
<td>0.94</td>
<td>0.72</td>
<td>0.68</td>
</tr>
</tbody>
</table>

**Directional Predictions**

- virus
- medication
- antioxidant

**Bi-Directional Predictions (synonyms)**

- ad agency
- bassist
- malware

- mosquito borne disease
- blood thinner
- fat soluble vitamin
- advertising agency
- bass player
- malicious software
Finding High-Quality Concepts

- Intuition: concepts with many strong signals (i.e., subsumption relations) are likely to be of high quality
- Compare
  - **Latte**: 50 most frequent concepts among top 1K predictions
  - Is-A Network (Similar Frequency): 50 concepts with similar frequencies of Is-A relations
  - Is-A Network (Most Frequent): 50 most frequent concepts
Concluding Observations

● Latte finds latent subsumptions that are useful for search engines using attributes
  ○ Is-A Network: 0.74 F1 / 2.4M predictions
  ○ Freebase: 0.98 F1 / 734 predictions
● Can also be used to clean resources with noisy concepts
● Challenge: What is the real notion of subsumption for Web applications?
Thanks!