Conceptual Similarity: Why, Where, How

Michalis Sfakakis

Laboratory on Digital Libraries & Electronic Publishing,
Department of Archives and Library Sciences,
Ionian University, Greece

First Workshop on Digital Information Management, March 30-31, 2011
Corfu, Greece

Do we need similarity?

- Are the following objects similar?
  - (Similarity, SIMILARITY)
    - As character sequences, NO!
      - How do they differ?
    - As character sequences, but case insensitive, Yes!
    - As English words, Yes!
      - Same word! They have the same definition, written differently

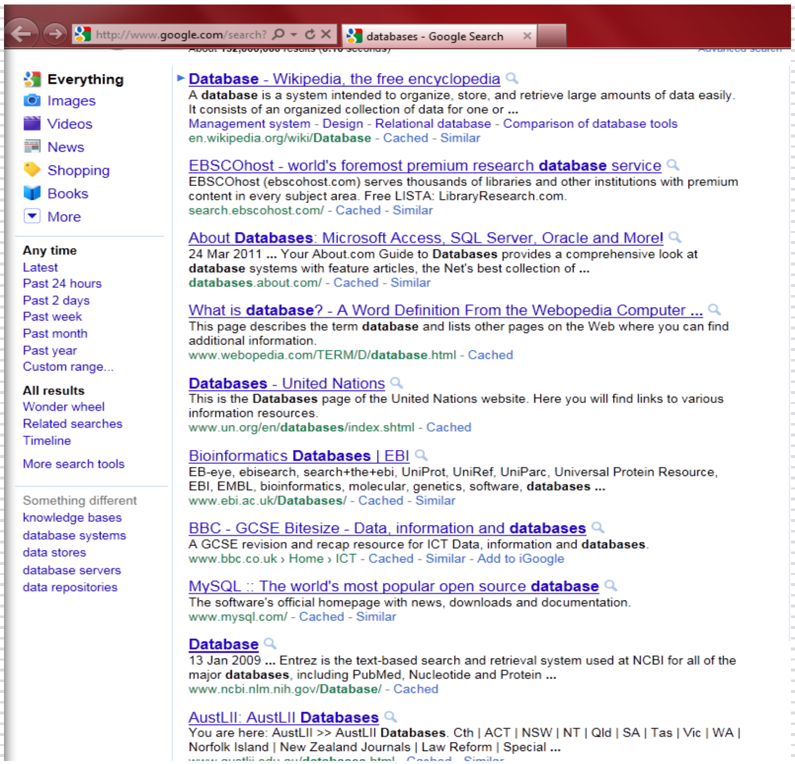
Contents

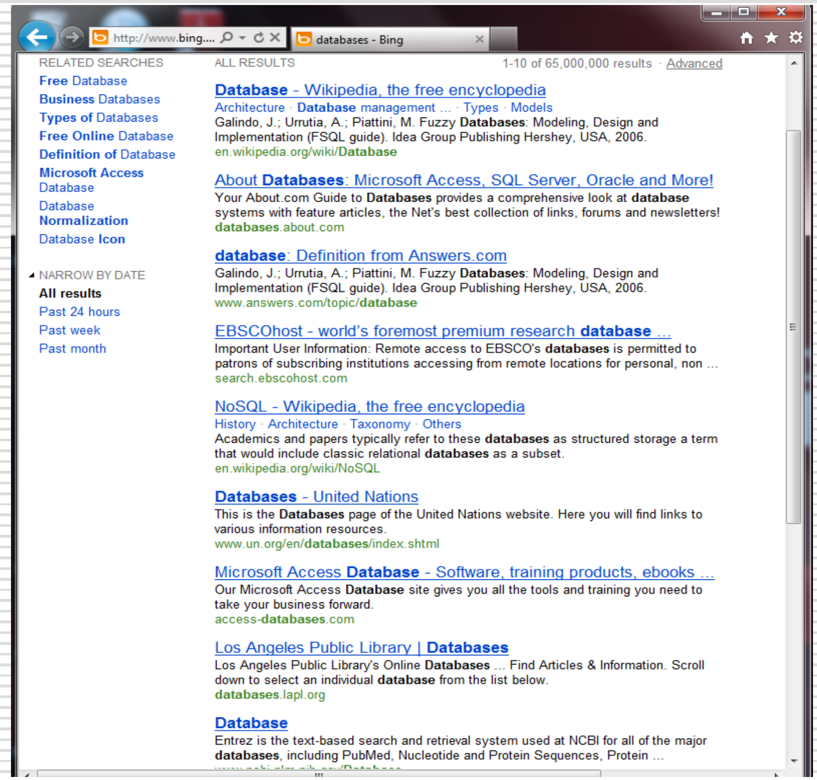
- Introduction
- Disciplines
- How we measure similarity
  - Focus on Ontology Learning evaluation

Exploring similarity… more cases

- What about the similarity of the objects?
  - (1, a)
    - The first object is the number one and the second is the first letter of the English alphabet. Therefore, as the first is a number and the second is a letter, they are different!
    - But, conceptually… When both represent an order, e.g. a chapter, or a paragraph number, they are both representing the first object of the list, the first chapter, paragraph, etc. Therefore, they could be considered as being similar!

Results for an Information Need



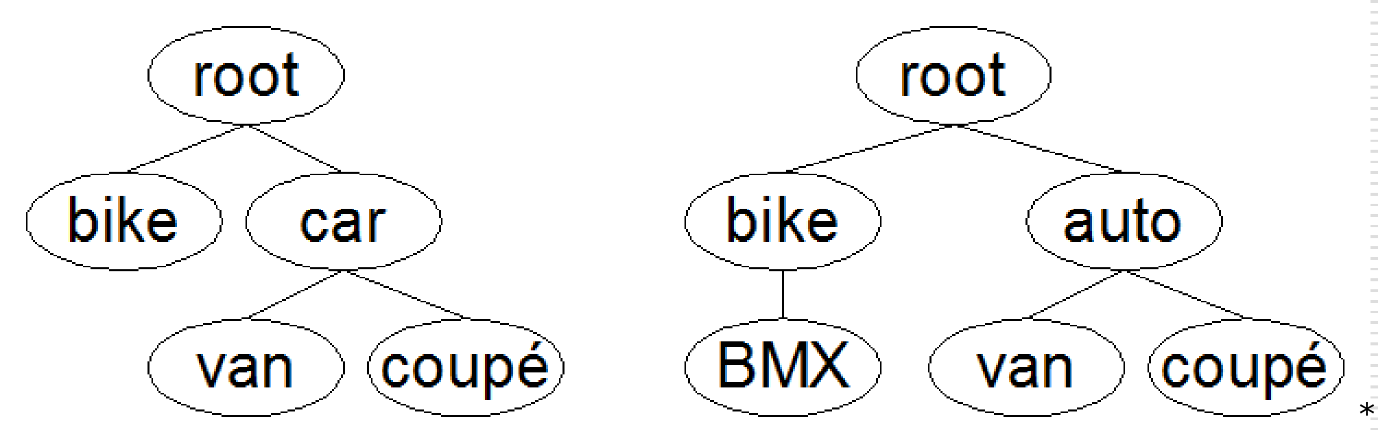


- How similar are the Results? Which one to select?

Comparing Concepts

- … again, how similar are the following objects?
  - (Disease, Illness)
    - As English words, or as character sequences they are not similar!
      - How do they differ?
    - As synonymous terms in a Thesaurus, they are both representing the same concept. (related with the *equivalency* relationship)

Comparing Hierarchies



*

- How similar…
  - … is the node *car* from the left hierarchy to the node *auto* from the right hierarchy?
  - … are the nodes *van* from both hierarchies?
  - … is the above hierarchies?

* [Dellschaft and Staab, 2006]

… so, what similarity is?

- Similarity is a context dependent concept
- Merriam-Webster's Learner's dictionary defines similarity as*:
  - A quality that makes one person or thing like another
  - … and similar, having characteristics in common
- Therefore, the context and the characteristics in common are required in order to specify and measure similarity

* http://www.learnersdictionary.com/search/similarity

Where the concept of similarity is encountered

- … Similarity is a context dependent concept

- Machine learning
  - Ontology Learning
  - Schema & Ontology Matching and Mapping
  - Clustering
  - IR
  - … in any evaluation concerning the results of a pattern recognition algorithm

- Vital part of the Semantic Web development

Precision & Recall in IR, measuring similarity between answers

- Let *C* be the result set for a query (the retrieved documents, i.e. the *Computed* set)
- Also, we need to know the correct results for the query (all the relevant documents, the *Reference* set)
  - *Precision*: is the fraction of retrieved documents that are relevant to the search
  - Recall: is the fraction of the documents that are relevant to the query that are successfully retrieved

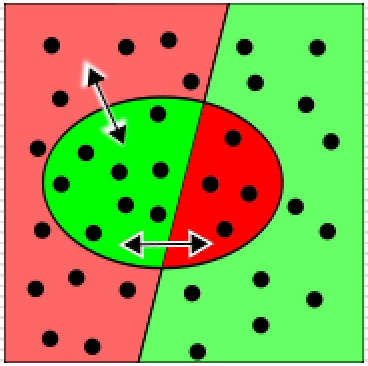
$$\text{precision} = \frac{|\{\text{relevant documents}\} \cap \{\text{retrieved documents}\}|}{|\{\text{retrieved documents}\}|} \qquad \text{recall} = \frac{|\{\text{relevant documents}\} \cap \{\text{retrieved documents}\}|}{|\{\text{relevant documents}\}|}$$

Wikipedia: http://en.wikipedia.org/wiki/Precision_and_recall

… Precision & Recall, a way to measure similarity

- *Precision* & *Recall* are two widely used metrics for evaluating the correctness of a pattern recognition algorithm

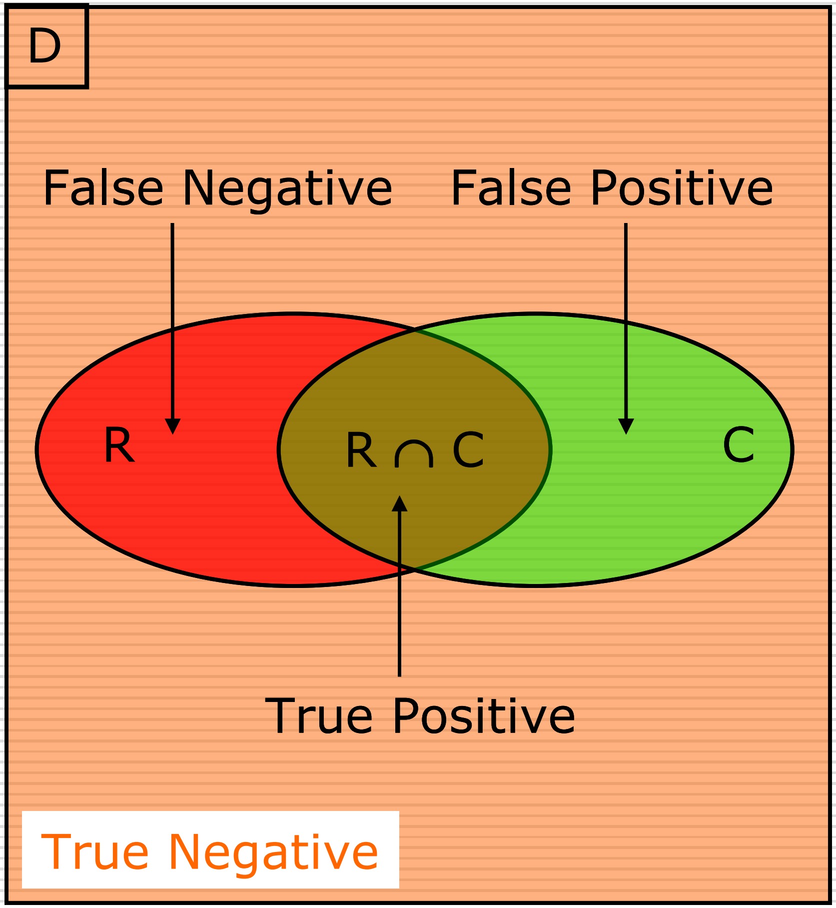
- *Recall* and *Precision* depend on the outcome (oval) of a pattern recognition algorithm and its relation to all relevant patterns (left) and the non-relevant patterns (right).
  The more correct results (green), the better.
  - *Precision*: horizontal arrow.
  - *Recall*: diagonal arrow.



Wikipedia: http://en.wikipedia.org/wiki/Precision_and_recall

Precision & Recall, once more

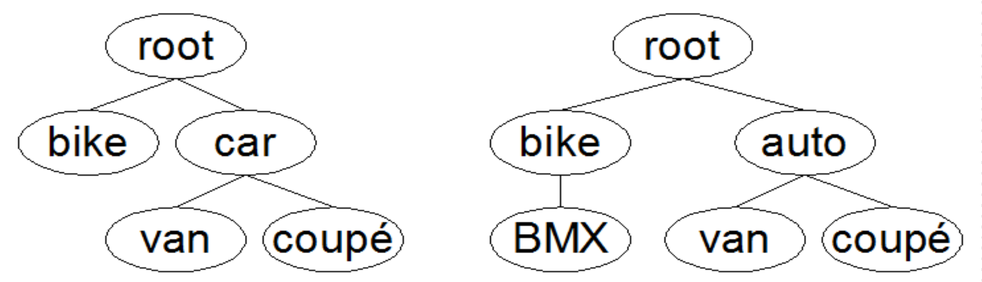
- Precision
  - $P = |R \cap C| / |R|$
- Recall
  - $R = |R \cap C| / |C|$
- $TP = R \cap C$
- $TN = D - (R \cup C)$
- $FN = R - C$
- $FP = C - R$



Overall evaluation, combining *Precision* & *Recall*

- Given *Precision* & *Recall*, *F-measure* could combines them for an overall evaluation
- Balanced *F-measure* (*P* & *R* are evenly weighted)
  - $F_1 = 2*(P*R)/(P+R)$
- Weighted *F-measure*
  - $F_b = (1+b^2)*(P*R)/(b^2*P+R)$, b non-zero
  - F_1 (b=2) weights recall twice as much as precision
  - $F_{0.5}$ (b=0.5) weights precision twice as much as recall

Measuring Similarity, Comparing two Ontologies

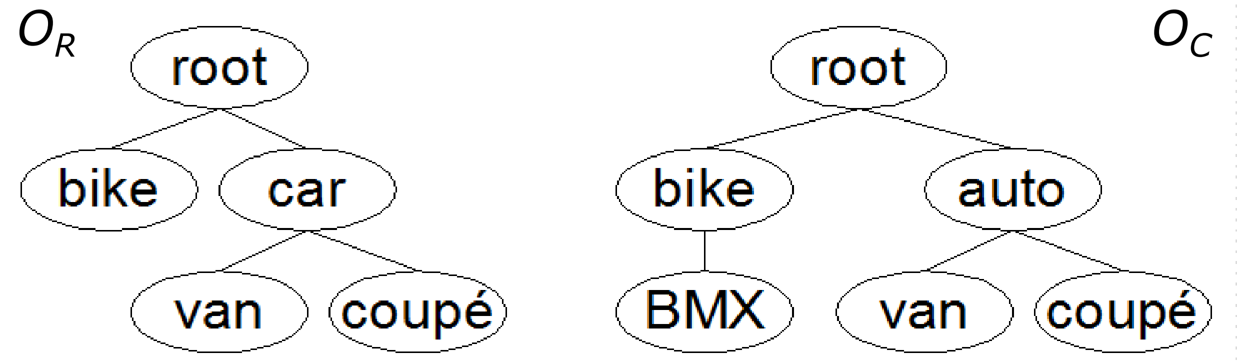


- A simplified definition of a core ontology*:
  - The structure $O := (C, root, \leq_C)$ is called a core ontology. C is a set of concept identifiers and *root* is a designated root concept for the partial order $\leq_C$ on C. This partial order is called concept hierarchy or taxonomy. The equation $\forall c \in C : c \leq_C root$ holds for this concept hierarchy.

- Levels of comparison
  - Lexical, how terms are used to convey meanings
  - Conceptual, which conceptual relations exist between terms
  - …

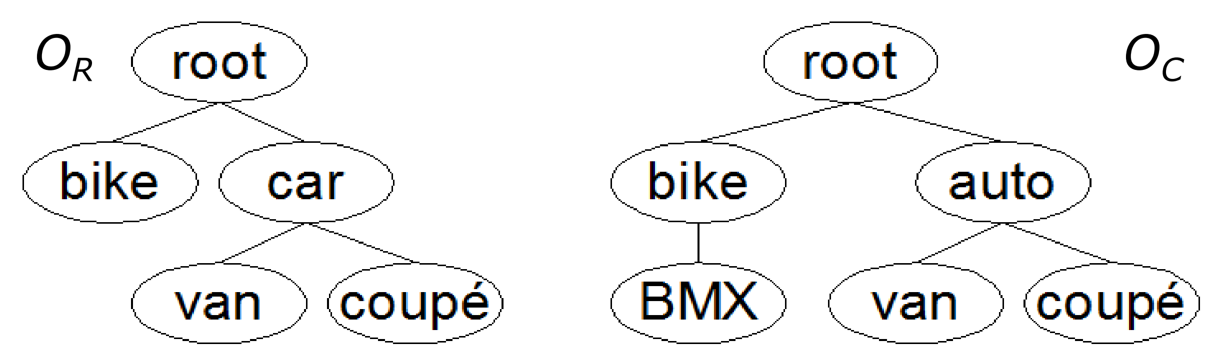
* [Dellschaft and Staab, 2006]

Gold Standard based Evaluation of Ontology Learning



- Given a pre-defined ontology
  - The so-called *Gold Standard* or *Reference*
- Compare the *Learned (Computed) Ontology* with the *Gold Standard*

Measuring Similarity - Lexical Comparison Level – LP, LR



- *Lexical Precision* & *Lexical Recall*
  - $LP(O_C, O_R) = |C_C \cap C_R| / |C_C|$
  - $LR(O_C, O_R) = |C_C \cap C_R| / |C_R|$

- The lexical precision and recall reflect how good the learned lexical terms C_C cover the target domain C_R
- For the above example LP=4/6=0.67, LR=4/5=0.8

Measuring Similarity, Lexical Comparison Level - aSM

- *Average String Matching*, using edit distance
  - *Levenshtein distance*, the most common definition for edit distance, measures the minimum number of token insertions, deletions and substitutions required to transform one string into an other
  - For example*, the *Levenshtein distance* between "*kitten*" and "*sitting*" is 3 (there is no way to do it with fewer than three edits)
    - **k**itten → **s**itten (substitution of 's' for 'k')
    - sitt**e**n → sitt**i**n (substitution of 'i' for 'e')
    - sittin → sittin**g** (insertion of 'g' at the end).

* Wikipedia: http://en.wikipedia.org/wiki/Levenshtein_distance

Measuring Similarity, Lexical Comparison Level – String Matching

- *String Matching* measure (SM), *given two lexical entries* L_1, L_2
  - Weights the number of the required changes against the shorter string
  - 1 stands for perfect match, 0 for bad match

$$\mathrm{SM}(L_i, L_j) := \max\left(0, \frac{\min(|L_i|, |L_j|) - \mathrm{ed}(L_i, L_j)}{\min(|L_i|, |L_j|)}\right) \in [0, 1]$$

- *Average SM*
  - *Asymmetric, determines the extend to which* $\mathcal{L}_1$ *(target) is covered by* $\mathcal{L}_2$ (source)

$$\overline{\mathrm{SM}}(\mathcal{L}_1, \mathcal{L}_2) := \frac{1}{|\mathcal{L}_1|} \sum_{L_i \in \mathcal{L}_1} \max_{L_j \in \mathcal{L}_2} \mathrm{SM}(L_i, L_j)$$

[Maedche and Staab, 2002]

Measuring Similarity, Lexical Comparison Level - *RelHit*

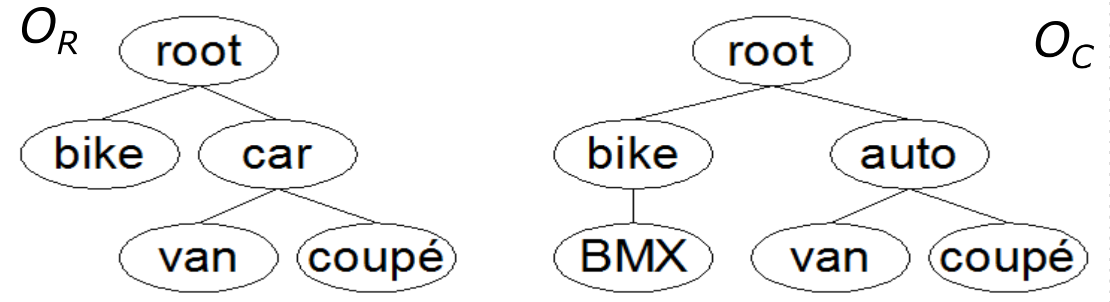
- Relative Number of Hits $$\mathrm{RelHit}(\mathcal{L}_1, \mathcal{L}_2) := \frac{|\mathcal{L}_1 \cap \mathcal{L}_2|}{|\mathcal{L}_1|}$$

- *RelHit* actually express Lexical Precision
- *RelHit* Compared to average String Matching
  - *Average SM* reduces the influences of string pseudo-differences (e.g. singular vs. plurals)
  - *Average SM* may introduce some kind of noise, e.g. “power”, “tower”

Measuring Similarity, Conceptual Comparison Level

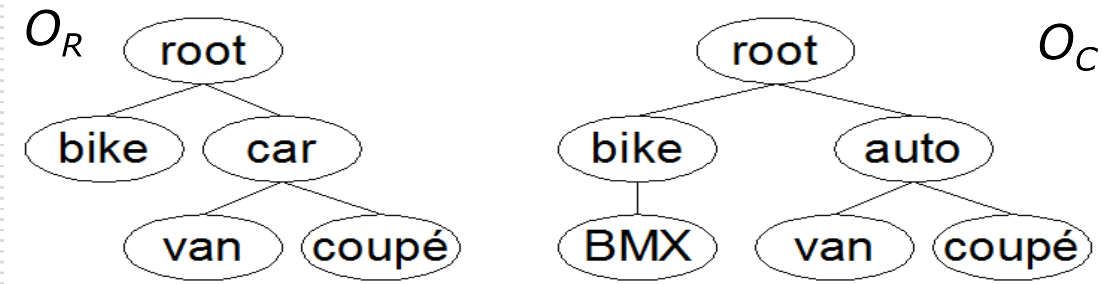
- Conceptual level compares semantic structure of ontologies
- Conceptual structures are constituted by Hierarchies, or by Relations
- How to compare two hierarchies?
- How do the positions of concepts influence similarity of Hierarchies?
- What measures to use?

Measuring Similarity, Conceptual Comparison Level



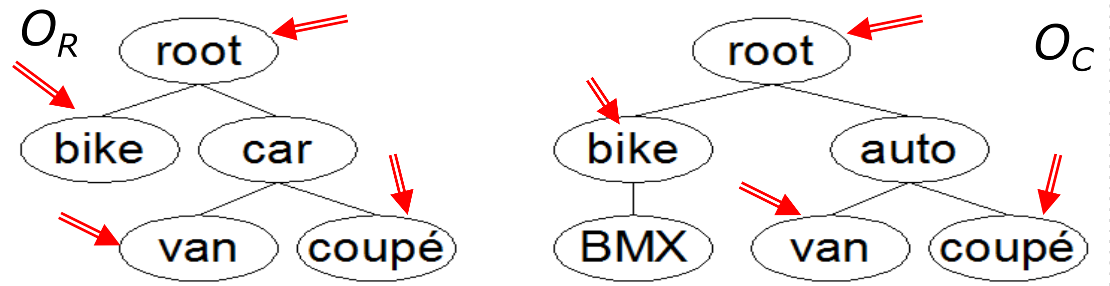
- Local measures compare the positions of two concepts based on characteristics extracts from the concept hierarchies they belong to
- Some characteristic extracts
  - Semantic Cotopy (sc)
    - $sc(c, O) = \{c_i | c_i \in C \wedge (c_i \leq c \vee c \leq c_i)\}$
  - Common Semantic Cotopy (csc)
    - $csc(c, O_1, O_2) = \{c_i | c_i \in C_1 \cap C_2 \wedge (c_i <_1 c \ \vee \ c <_1 c_i)\}$

Measuring Similarity, Conceptual Comparison Level – sc



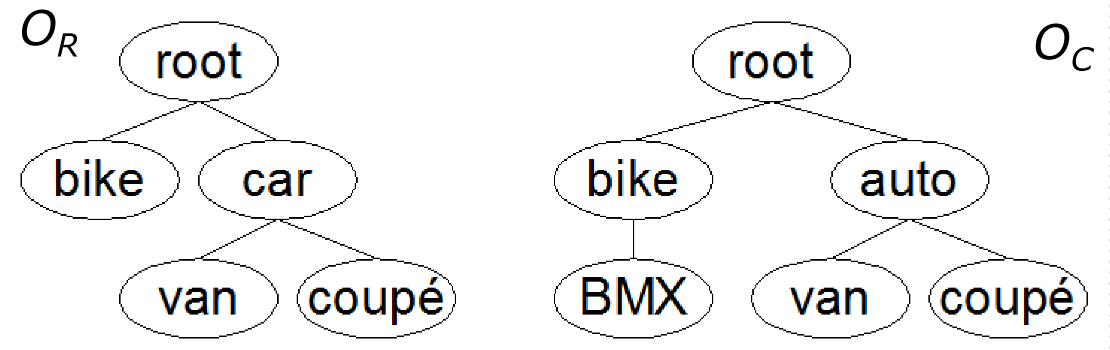
- Semantic Cotopy
  - $sc(c, O) = \{c_i | c_i \in C \wedge (c_i \leq c \vee c \leq c_i)\}$
- Semantic Cotopy examples
  - sc("root", O_R) = {root, bike, car, van, coupé}
  - sc("root", O_C) = {root, bike, auto, BMX, van, coupé}
  - sc("bike", O_R) = {root, bike}
  - sc("bike", O_C) = {root , bike, BMX}
  - sc("car", O_R) = {root , car, van, coupé}
  - sc("auto", O_C) = {root, auto, van, coupé}

Measuring Similarity, Conceptual Comparison Level – csc



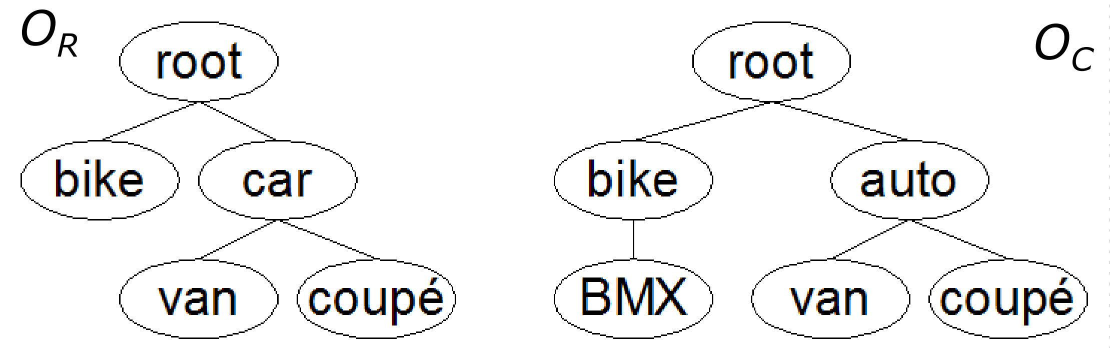
- Common Semantic Cotopy
  - $csc(c, O_1, O_2) = \{c_i | c_i \in C_1 \cap C_2 \wedge (c_i <_1 c \ \vee \ c <_1 c_i)\}$
- Common Semantic Cotopy examples
  - $C_1 \cap C_2$ = {root, bike, van, coupé}
  - csc("root", o_R, o_C) = {bike, van, coupé}
  - csc("root", o_C, o_R) = {bike, van, coupé}
  - csc("bike", o_R, o_C) = {root}, csc("bike", o_C, o_R) = {root}
  - csc("car", o_R, o_C) = {root , van, coupé}, csc("car", o_C, o_R) = $\varnothing$
  - csc("auto", o_C, o_R) = {root, van, coupé} }, csc("auto", o_C, o_R) = $\varnothing$

Measuring Similarity, Conceptual Comparison Level – local measures tp, tr



- Local *taxonomic precision* using characteristic extracts
  - $tp_{ce}(c_1, c_2, O_C, O_R) = |ce(c_1, O_C) \cap ce(c_1, O_R)| / |ce(c_1, O_C)|$
- Local *taxonomic recall* using characteristic extracts
  - $tr_{ce}(c_1, c_2, O_C, O_R) = |ce(c_1, O_C) \cap ce(c_1, O_R)| / |ce(c_1, O_R)|$

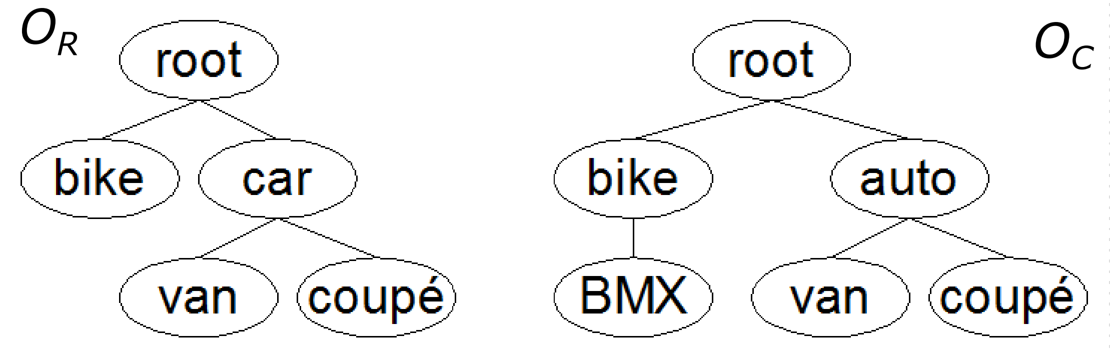
Measuring Similarity, Conceptual Comparison Level – local measures tp



- *Local taxonomic precision* examples using *sc*
  - sc("bike", O_R) = {root, bike},
    sc("bike", O_C) = {root, bike, BMX}
  - tp_{sc}("bike", "bike", O_C, O_R) = |{root, bike}|/|{root, bike, BMX}|,
    tp_{sc}("bike", "bike", O_C, O_R) = 2/3 = 0.67

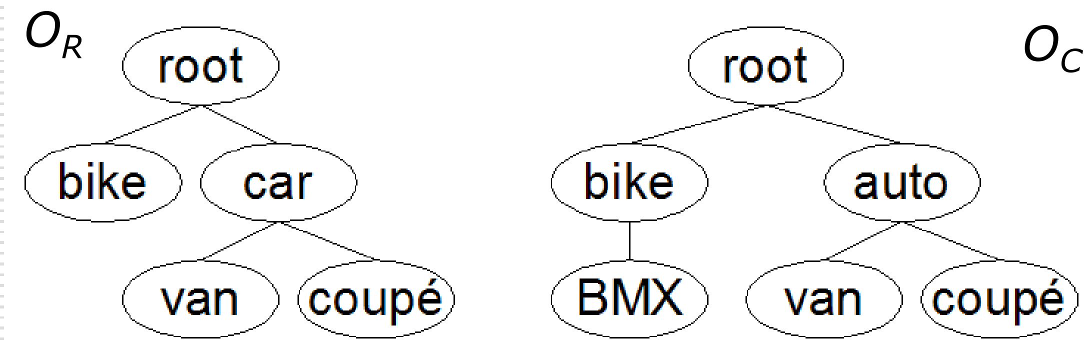
[Maedche and Staab, 2002]

Measuring Similarity, Conceptual Comparison Level – local measures tp



- *Local taxonomic precision* examples using *sc*
  - sc(“car”, O_R) = {root , car, van, coupé},
    sc(“auto”, O_C) = {root , auto, van, coupé}
  - *tp_{sc}(“car”, “auto”, O_C, O_R) =*
    |{root, van, coupé} |/|{root, auto, van, coupé}|,
    tp_{sc}(“car”, “auto”, O_C, O_R) = 3/4 = 0.75

Measuring Similarity, Conceptual Comparison Level – comparing Hierarchies



- *Global Taxonomic Precision (TP)*

$$TP(\mathcal{O}_C,\mathcal{O}_R) := \frac{1}{|\mathcal{C}_C|}\sum_{\underbrace{\boxed{c\in\mathcal{C}_C}}_{\text{concept set}}} \begin{cases} \overbrace{\boxed{tp(c,c,\mathcal{O}_C,\mathcal{O}_R) \quad \text{if } c\in\mathcal{C}_R}}^{\text{local taxonomic precision}} \\ \underbrace{\boxed{\max_{c'\notin\mathcal{C}_R} tp(c,c',\mathcal{O}_C,\mathcal{O}_R) \quad \text{if } c\notin\mathcal{C}_R}}_{\text{estimation}} \end{cases}$$

Measuring Similarity, Conceptual Comparison Level – Overall evaluation

- … again *F-measure*, but now using *Global Taxonomic Precision (TP)* and *Global Taxonomic Recall (TR)*
- Balanced Taxonomic *F-measure* (T*P* & T*R* are evenly weighted)
  - $TF_1 = 2*(TP*TR)/(TP+TR)$
- Weighted T*F-measure*
  - $TF_b = (1+b^2)*(TP*TR)/(b^2*TP+TR)$, b non-zero

  - TF_1 (b=2) weights recall twice as much as precision
  - $TF_{0.5}$ (b=0.5) weights precision twice as much as recall

Measuring Similarity, Conceptual Comparison Level – Taxonomic Overlap

- *Global Taxonomic Overlap… based on local taxonomic overlap (TO)*

$$\overline{TO}(O_1, O_2) = \frac{1}{|C_1|} \sum_{c \in C_1} TO(c, O_1, O_2)$$

$$TO(c, O_1, O_2) = \begin{cases} TO'(c, O_1, O_2) \; if \; c \in C_2 \\ TO''(c, O_1, O_2) \; if \; c \notin C_2 \end{cases}$$

$$TO'(c, O_1, O_2) := \frac{|SC(c, O_1, O_2) \cap SC(c, O_2, O_1)|}{|SC(c, O_1, O_2) \cup SC(c, O_2, O_1)|}$$

$$TO''(c, O_1, O_2) := \max_{c' \notin C_2} \frac{|SC(c, O_1, O_2) \cap SC(c', O_2, O_1)|}{|SC(c, O_1, O_2) \cup SC(c', O_2, O_1)|}$$

References & Further Reading

- Dellschaft, Klaas and Staab, Steffen (2006) : On How to Perform a Gold Standard Based Evaluation of Ontology Learning. In: I. Cruz et al. (Eds.) ISWC 2006. LNCS 4273, pp. 228–241. Springer, Heidelberg
- Maedche, A., Staab, S. (2002): Measuring similarity between ontologies. In: Proceedings of the European Conference on Knowledge Acquisition and Management (EKAW-2002). Siguenza, Spain
- Staab, S. and Hotho, A.: Semantic Web and Machine Learning Tutorial (available at http://www.uni-koblenz.de/~staab/Research/Events/ICML05tutorial/icml05tutorial.pdf)

- Bellahsene, Z., Bonifati, A., Rahm, E. (Eds.) (2011): Schema Matching and Mapping, Heidelberg, Springer- Verlag, ISBN 978-3-642-16517-7.

End of tutorial!

- Thanks for your attention!
  - Michalis Sfakakis