



# PLIN037 Sémantika a počítače

Zuzana Nevěřilová  
2020/21

# Znalost světa

Implicitní → Explicitní znalost

Objekty:

- Popis typického představitele

Situace:

- Typické situace  
(valenční rámce)

Encyklopedická znalost

- Německo leží v Evropě

Common sense:

- Co je větší než myš?
- Způsobí rána knihou do hlavy smrt?
- Kde se běžně nachází jídlo?

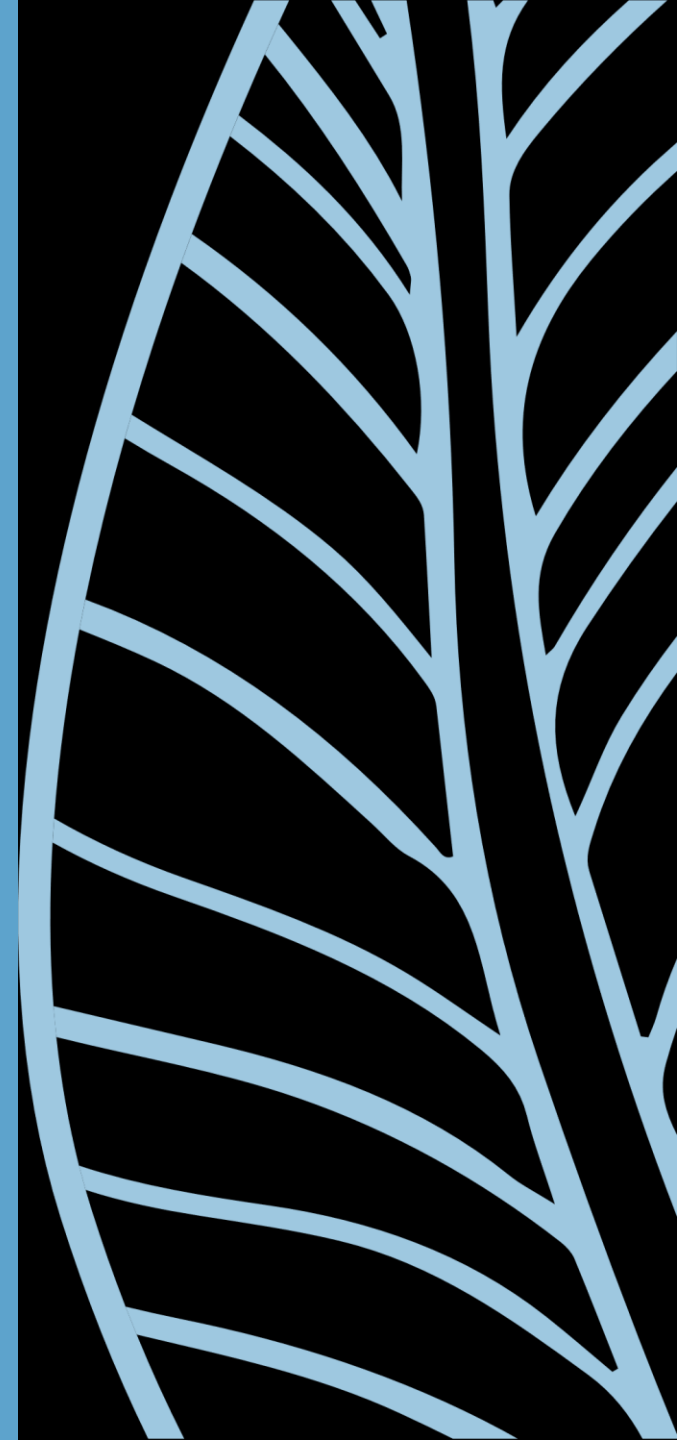
# Ontologie – sémantické sítě

Ontologie = explicitní specifikace sdílené konceptualizace (Gruber, 1995)

**Explicitní: co je řečeno, platí. Co není řečeno, neplatí.**

**Sdílené: lidé by se na tom shodli.**

**Konceptualizace: entity existují a existují i relace mezi nimi**



# Sémantická síť a rel

- nadtyp–podtyp, is a, is-a, isa (hypo/hyperonymie)
  - instance třídy, member of
  - část–celek, has a (holo/meronymie)
  - umístění
- 
- opozitnost (antonymie)
  - příčina–následek
  - ...

[https://cs.m.wikipedia.org/wiki/Soubor:Semantic\\_Net.svg](https://cs.m.wikipedia.org/wiki/Soubor:Semantic_Net.svg)

# Taxonomie

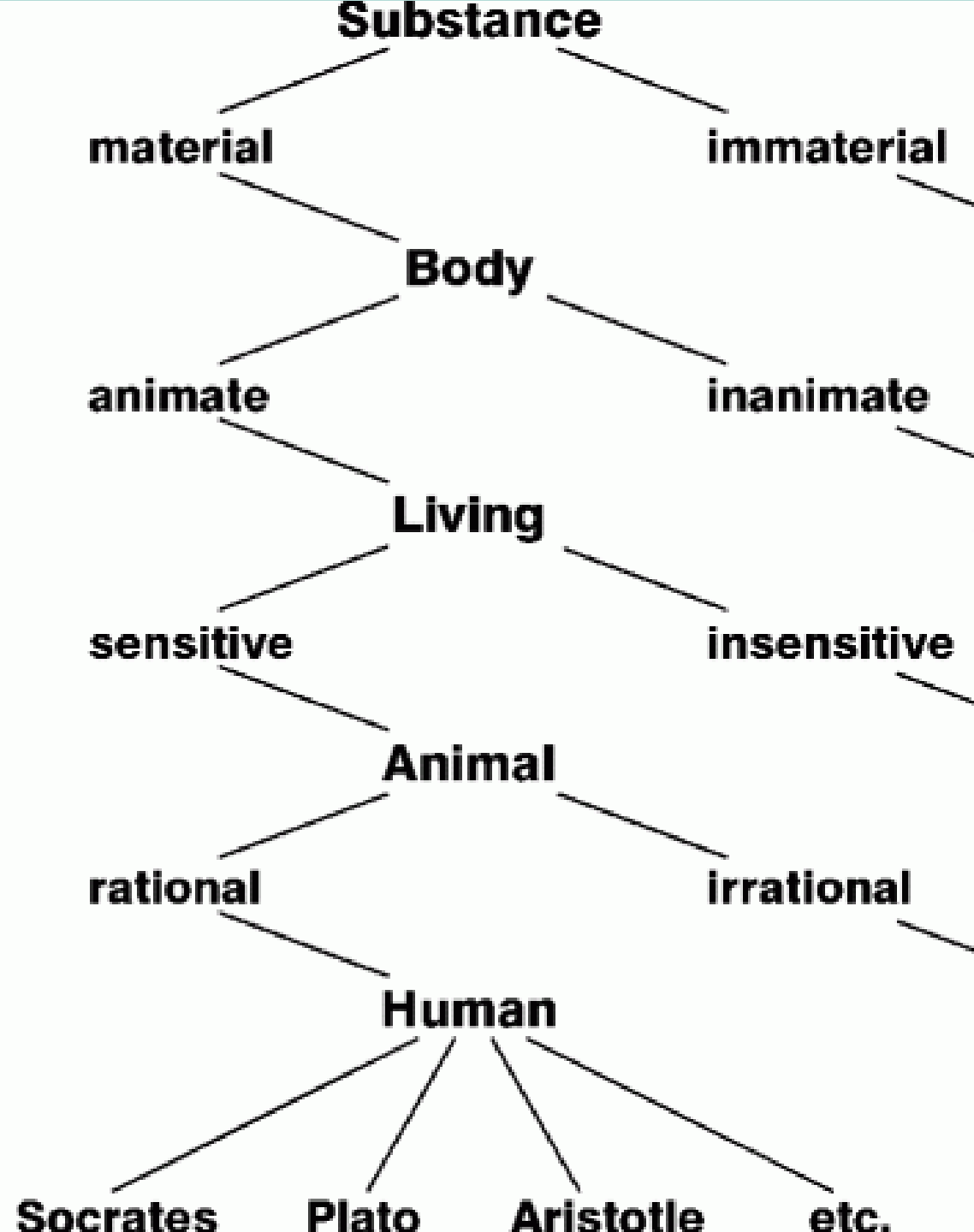
Aristoteles – kategorie (všech) entit, které mohou lidé vnímat

Porfyrios – uspořádal kategorie

Carl Linné – klasifikace (všech) organismů

Důležité rysy:

- uzly jsou třídy (organismů, entit . . . )
- koncové uzly jsou individua (instance)
- třídy jsou strukturované do stromu (podtřída, nadtřída)
- uzly na stejné úrovni se vzájemně vylučují (implicitní předpoklad)



# Odvozování ve znalostních bázích

Fakt  $F$  = tvrzení s pravdivostní hodnotou (např. ptáci létají)

Báze znalostí (knowledge base)  $KB$  = (pokud možno konzistentní) soubor faktů (např. ptáci létají, vlaštovka je pták)

Pokud z  $KB$  plyne  $F$  a přidáme další fakt takový, že  $KB$  je stále konzistentní, je  $KB$  **monotónní** reprezentace. (Allen, 1995)

ptáci létají  
vlaštovka je pták

---

vlaštovka létá

ptáci létají  
tučňák je pták

---

tučňák létá

ptáci kromě tučňáků létají  
tučňák je pták

---

# Odvozování ve znalostních bázích

V přirozeném jazyce potřebujeme **nemonotónní odvozování**.

Použijeme tzv. **implicitní pravidlo (default rule)**, které platí vždy, když neexistuje sporné pravidlo pro podtřídu.

**Ptáci létají. Tučňák je pták. Tučňák nelétá.**

ptáci kromě tučňáků, mlád'at, pštrosů, kiwi, mrtvých ptáků atd. létají

X je pták (kromě tučňáků, mlád'at, pštrosů, kiwi, mrtvých ptáků atd. létají)

---

X létá

# Existující sémantické sítě

## Common sense

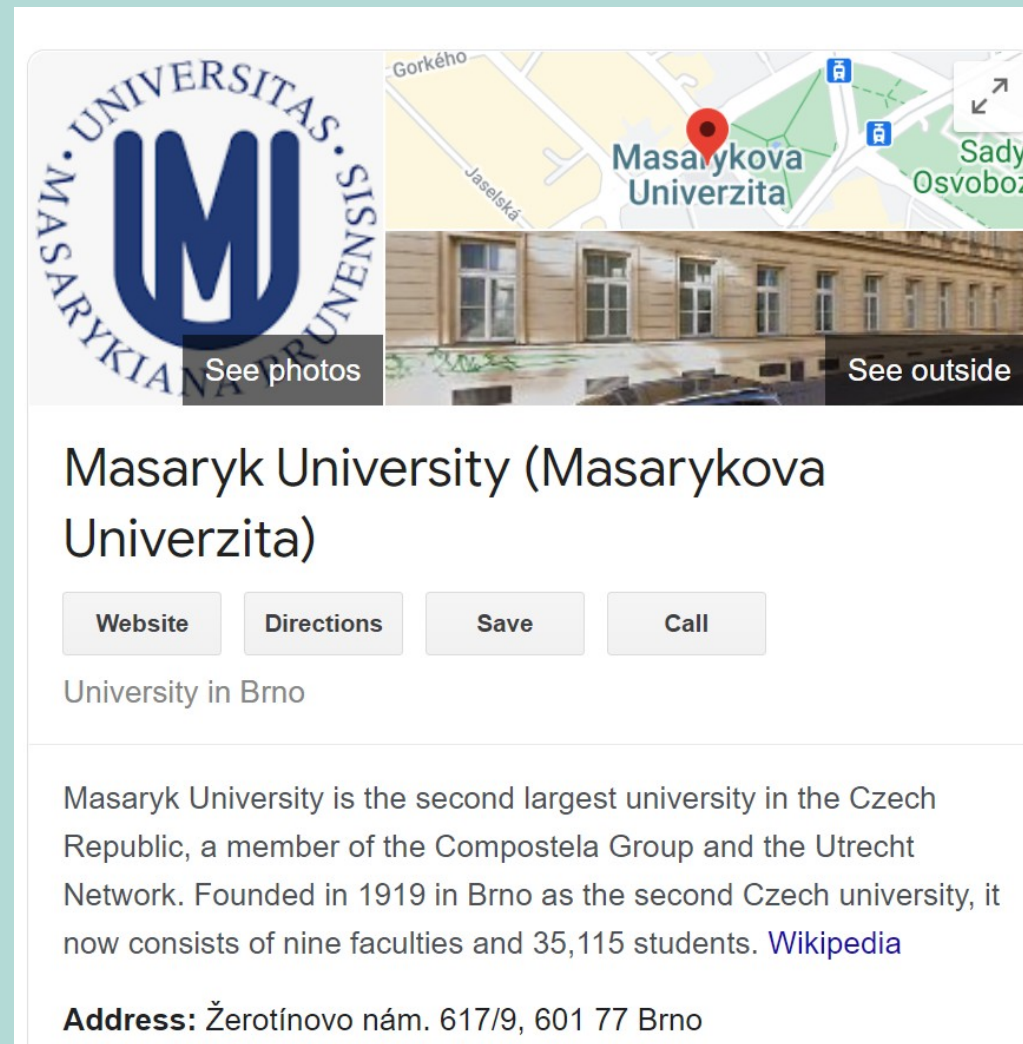
- ConceptNet
- Microsoft Concept Graph

## Lexikální sítě

- WordNet
- Wiktionary

## (Encyklopedické) znalosti

- YAGO
- SUMO/MILO
- dbPedia
- Google Knowledge Graph / schema.org



**Masaryk University (Masarykova Univerzita)**

Website Directions Save Call

University in Brno

Masaryk University is the second largest university in the Czech Republic, a member of the Compostela Group and the Utrecht Network. Founded in 1919 in Brno as the second Czech university, it now consists of nine faculties and 35,115 students. [Wikipedia](#)

**Address:** Žerotínovo nám. 617/9, 601 77 Brno



# WordNet

Původně Princeton WordNet (PWN) pro angličtinu: 117 tisíc synsetů

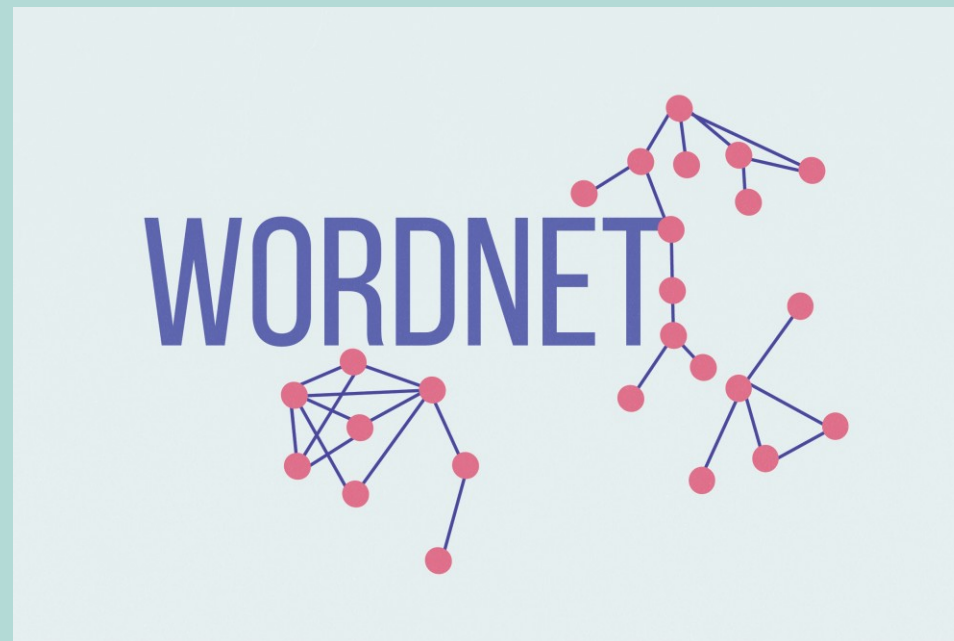
EuroWordNet – evropské jazyky, vrcholová ontologie (Top Ontology), základní koncepty (Base Concepts), interlingual identifier (ILI)

BalkaNet – balkánské a evropské jazyky

Global WordNet Association – sdružuje WordNety pro všechny jazyky

Český WordNet (FI MUNI) – 28 tisíc synsetů, automaticky rozšířená verze: 40 tisíc synsetů

Největší WordNet – polská Słowosieć: 294 tisíc synsetů



# Obecné a doménové ontologie

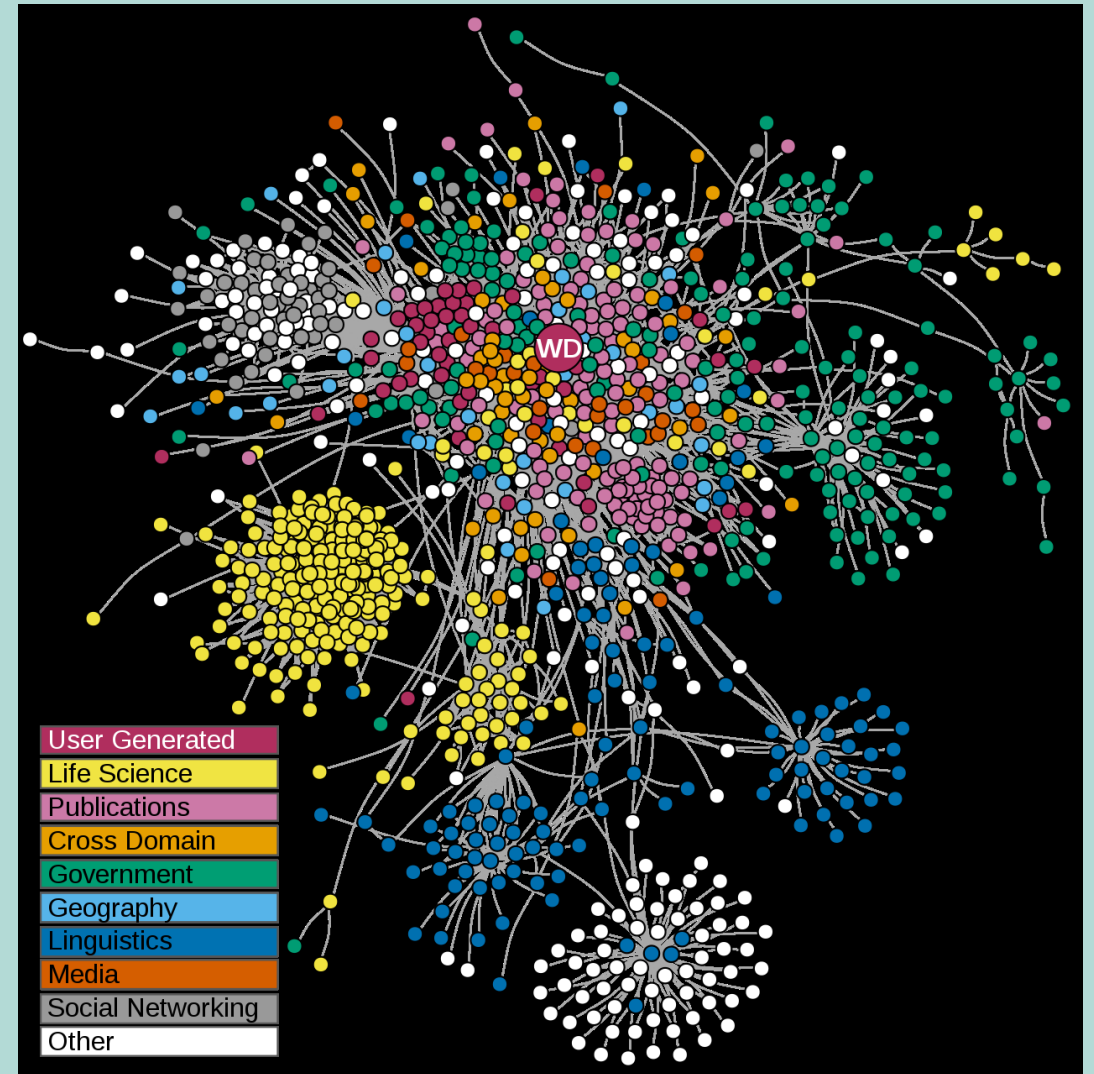
SUMO/MILO – Suggested Upper Merged Ontology/Mid-Level Ontology

YAGO – Yet Another Great Ontology

NELL – Never-Ending Language Learning

Doménové ontologie: GeoNames, Wikidata, Global Research Identifier Database, ...

## Linked (Open) Data



# Wikipedia a dbPedia

- (Polo)automaticky vygenerovaná znalostní báze
- Základem jsou tzv. Infoboxy z Wikipedie, kategorizace článků, obrázky, geografické souřadnice a externí odkazy
- Výhodou je, pokud wikipedista používá šablony stránek
- Poskytuje rozhraní přes Virtuoso SPARQL, kde je možné zadávat komplexní dotazy
- Obsahuje data z více než 100 Wikipedií

## Římskokatolická farnost Těšetice u Olomouce



farní kostel

### *Základní údaje*

<b>Děkanát</b>	Olomouc
<b>Diecéze</b>	arcidiecéze olomoucká
<b>Provincie</b>	moravská
<b>Farář</b>	R. D. Mgr. Mirosław Łukasiewicz

### *Území farnosti*



# Literatura

- Gruber, Thomas. Toward Principles for the Design of Ontologies Used for Knowledge Sharing. *International Journal Human-Computer Studies* Vol. 43, Issues 5-6, November 1995, p.907-928.  
<https://tomgruber.org/writing/onto-design.pdf>
- Gruber, Thomas. Ontology. In Liu, L. and Özsu, M. T., editors, *Encyclopedia of Database Systems*, Springer Verlag. 2009. pp 1963–1965. [https://link-springer-com-443.webvpn.zisu.edu.cn/referenceworkentry/10.1007/978-1-4614-8265-9\\_1318](https://link-springer-com-443.webvpn.zisu.edu.cn/referenceworkentry/10.1007/978-1-4614-8265-9_1318)
- Liu, H., Singh, P. ConceptNet — A Practical Commonsense Reasoning Tool-Kit. *BT Technology Journal* 22, 211–226 (2004).  
<https://doi.org/10.1023/B:BTTJ.0000047600.45421.6d>  
<http://alumni.media.mit.edu/~hugo/publications/papers/BTTJ-ConceptNet.pdf>
- Fellbaum, Christiane: *WordNet: An Electronic Lexical Database*. Bradford Books. 1998.
- Allen, J. *Natural Language Understanding (2nd ed.)*. Benjamin-Cummings Publishing Co., Inc., Redwood City, CA, USA. 1995