



PLIN021 SÉMANTICKÁ ANALÝZA V PRAXI

ZUZANA NEVĚŘILOVÁ

2020/21

LEXIKÁLNÍ DESAMBIGUACE

OPAKOVÁNÍ

- Lexikální desambiguace je přiřazení slova v daném kontextu významu v určitém repozitáři významů (např. Slovníku).
- Leskův algoritmus řeší úlohu porovnáním kontextu slova se slovy z definice významu.
- V definici nutně nemusejí být slova, která se vůbec kdy s hledaným slovem vyskytují.
- Úspěch algoritmu silně závisí na použitém slovníku.
- Slovníky nebyly napsány s cílem být zdrojem pro algoritmus WSD.
- Je tedy třeba se porozhlédnout po **jiných algoritmech**.

LEXIKÁLNÍ DESAMBIGUACE (YAROWSKI, 1995)

hledáme význam s slova w

1. vezmi všechny výskyty slova w z korpusu včetně jejich **kontextů**
2. pro každý možný **význam** slova, vytvoř malou sadu příkladů (buď ručně, nebo pomocí kolokací)
3. vytvoř **rozhodovací seznam** s pravděpodobnostmi pro další slova, která se vyskytují v kontextech a aplikuj tento seznam na celý korpus (s prahem pro pravděpodobnost)
4. nově zařazená slova obsahují **další slova** v kontextech
5. algoritmus můžeme upravit pomocí zařazení předpokladu *one-sense-per-discourse*
6. opakuj kroky 3–6
7. Jakmile množiny označených výskytů přestanou narůstat, zastav
8. systém je nyní natrénovaný i na jiný korpus!

LEXIKÁLNÍ DESAMBIGUACE (YAROWSKI, 1995)

1. vezml všechny výskyty slova *w* z korpusu včetně jejich kontextů

místo – místo benzínovou pumpu, jestli ty **kočky** nepřemístili jinde. "Nevíte, kde jsou ty **kočky** , co měly být na výstavišti?" "Kočky? Zeptejte ty kočky, co měly být na výstavišti?" " **Kočky** ? Zeptejte se kolegyně, já nedávno nastoupila ale nevím. U nás žádný plakát nevisel." " **Kočky** se nevedou!" povzdechla jsem a zavrtěla zaslechl a přidal se k hovoru. "Chcete vidět **kočky** ?" "Jste místní? Ano, chceme! Tam nahoře plešatých, mouratých, černých, bílých a jiných **koček** . "Chudinky, trápí je," ozvala jsem se. se sud' te! A ostatní se přidávali, brali **kočky** pro sebe i pro sousedy, pro známé a příbuzné No a pan z posledního auta si vzal místo **kočky** překvapenou Jarunu a odvezl ji k sobě do " Jindy se s ní zkusím domluvit: "Hele, **Kočko** , dneska by se mi vážně hodilo, abys zůstala aby to dneska všechno dobře dopadlo..." **Kočka** je Osobnost a kašle na můj osud. Trénuje Postupem času zjišťuji, že jsem se naučila brát **Kočku** jako zvířecího kamaráda, který nemá s mým tlapkách moji budoucnost ani věci příští. **Kočka** je zkratka Kočka. Ulevilo se mi tímto zjištěním budoucnost ani věci příští. Kočka je zkratka **Kočka** . Ulevilo se mi tímto zjištěním. Cítím se V některých domácnostech spolu kamarádí **kočka** a pes. V jiných zase spolu žijí masochistka volný čas, energii a peníze na záchranu psů, **koček** a dalších zvířat. Proto tak obdivuji a

LEXIKÁLNÍ DESAMBIGUACE (YAROWSKI, 1995)

2. pro každý možný význam slova, vytvoř malou sadu příkladů (buď ručně, nebo pomocí kolokací)

Jenže - pokud si lidé budou hady plést s	kočkami nebo psy	, může docházet k úrazům a na to je třeba
vychytralou mouchou, která narazila na sklo.	Kočka vrtela vousama. pes	byl mrtvej že žrása okupoval stránky pro
bylí uvěznění osadníci nucení jíst koně,	kočky, psy	a dokonce i krysy, které nechtěně přivezli
chov. </p><p> Kočička do krabice </p><p> Znáte	kočky? Kočka není pes	, mazlí se jen, když sama chce, v opačném
<p> Kočička do krabice </p><p> Znáte kočky?	Kočka není pes	, mazlí se jen, když sama chce, v opačném
koťátkem bude jen pár měsíců. Snese se s	kočkou váš pes	? Víte, co kočka potřebuje? Než si kotě
onemocnění, jako je chřipka či nevolnost. Ovšem u	koček, psů	, kuřat či prasat koronaviry způsobují vážná
2005) </p><p> Možná jste si jako majitelé	koček a psů	někdy uvědomili rozdíl mezi těmito domácími
Pop Artové prvky a Warhol začal malovat	kočky, psy	a prasata, jež byly oblíbeným Wyethovým
jsem někoho v baráku vzbudil. Vlastně jo -	kočky, pes	Alfík i bejček Kryštof už čekají v pozoru
dentifikačního mikročipu používán již delší dobu na	kočkách, psech	a koních: na Novém Zélandu již navíc na
pokračují snahy o vyhubení potkanů, zdivočelých	koček, psů	, prasat a koz na ostrovech. </p><p> Kde je
stádo uzavírá. Při nebezpečí (cizí člověk,	kočka, pes), Slasy "sykají" a podupávají předními
Smrtelná dávka pro krysu je asi 0,02 g, pro	kočky a psy	je smrtelná dávka asi 25krát vyšší. </p>
V některých domácnostech spolu kamarádí	kočka a pes	. V jiných zase spolu žijí masochistka se

LEXIKÁLNÍ DESAMBIGUACE (YAROWSKI, 1995)

2. pro každý možný **význam** slova, vytvoř malou sadu příkladů (buď ručně, nebo pomocí kolokací)

hned zaujala, byla to typická prvotřídní	kočka, blond vlasy	, pevná větší prsa natěsnané v přiléhavém
okouzlí nejen španělská korida, ale i lokální	kočka s havraními vlasy	, jiskrou v těle a jménem Adela v rodném
) </p><p> 24.01.2008 polypocket11: Bálová	kočka:zrzavý vlasy	, hnědý oči, žlutý stíny, úplně poslední rtěnku
????? </p><p> 02.01.2008 Murtagh1: PLESOVÁ	KOČKA: Blond roztřepeně vlasy	, zelené oči, modré stíny, černé šaty, ty
sklopec jsme vyzkoušeli, bohužel také jen	kočky.vlasy	jsme vyzkoušeli a lezou na půdu dál. koupili
prochází další lidé a každému dvě pohledné	kočky strkají do vlasů	nějaké drátky. Ihned jsem chtěl vědět jak
dětství Markétku. Stala se z ní pěkná vospělá	kočka (dlouhé černé vlasy	, štihlá postava, prsa č. 3, vypasované

LEXIKÁLNÍ DESAMBIGUACE (YAROWSKI, 1995)

3. vytvoř **rozhodovací seznam** s pravděpodobnostmi pro další slova, která se vyskytují v kontextech a aplikuj tento seznam na celý korpus (s prahem pro pravděpodobnost)

```
if ('pes' in context(w)) then s(w,A)=1
if ('vlasy' in context(w)) then s(w,B)=1
if ('prase' in context(w)) then s(w,A)=0.9
if ('prsa' in context(w)) then s(w,B)=1
if ('oči' in context(w)) then s(w,B)=0.6
if ('člověk' in context(w)) then s(w,A)=0.8
```

LEXIKÁLNÍ DESAMBIGUACE (YAROWSKI, 1995)

4. nově zařazená slova obsahují **další slova** v kontextech

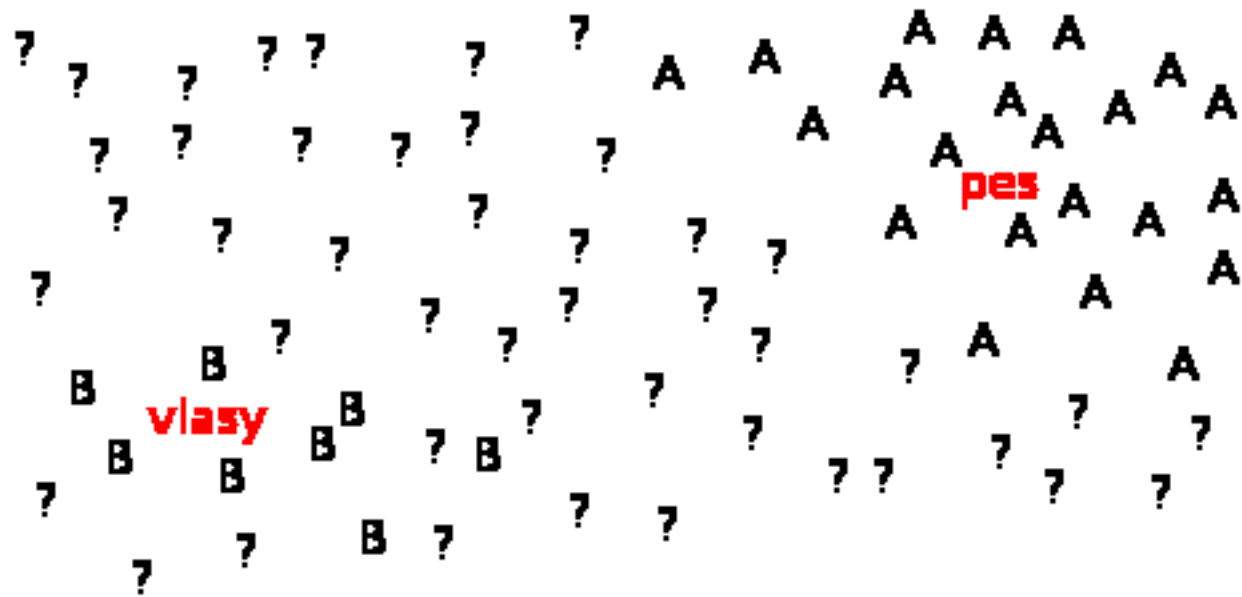
pes → majitel, žrát
vlasy → pohledná

LEXIKÁLNÍ DESAMBIGUACE (YAROWSKI, 1995)

hledáme význam s slova w

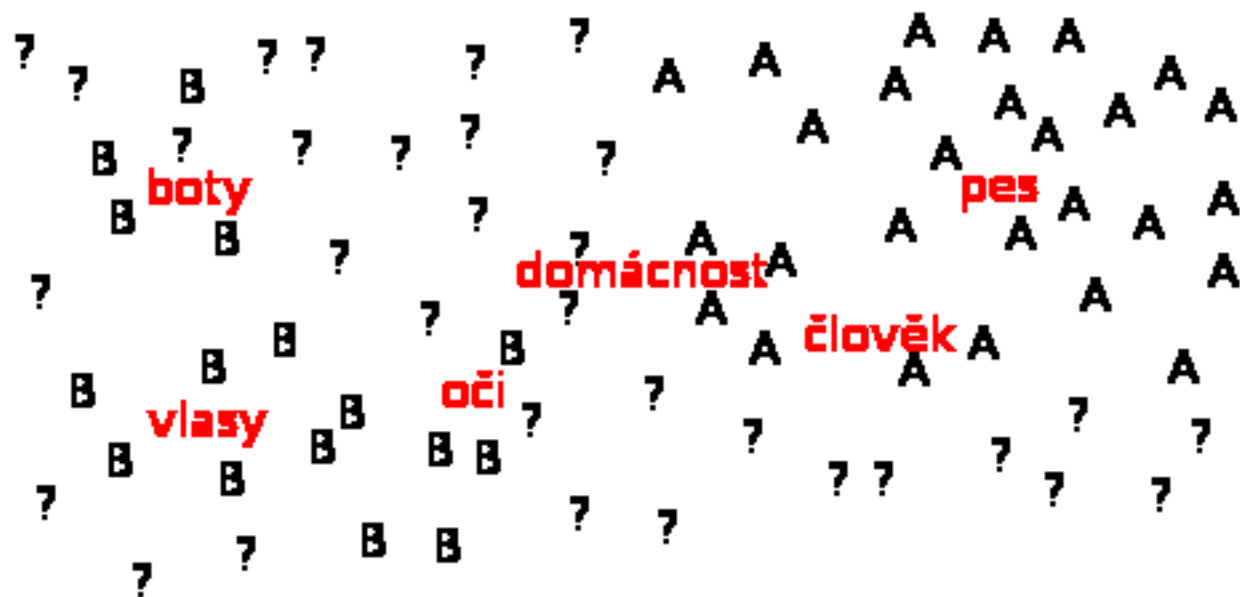
1. vezmi všechny výskyty slova w z korpusu včetně jejich **kontextů**
2. pro každý možný **význam** slova, vytvoř malou sadu příkladů (buď ručně, nebo pomocí kolokací)
3. vytvoř **rozhodovací seznam** s pravděpodobnostmi pro další slova, která se vyskytují v kontextech a aplikuj tento seznam na celý korpus (s prahem pro pravděpodobnost)
4. nově zařazená slova obsahují **další slova** v kontextech
5. algoritmus můžeme upravit pomocí zařazení předpokladu *one-sense-per-discourse*
6. opakuj kroky 3–6
7. Jakmile množiny přestanou narůstat, zastav
8. systém je nyní natrénovaný i na jiný korpus!

LEXIKÁLNÍ DESAMBIGUACE (YAROWSKI, 1995)



LEXIKÁLNÍ DESAMBIGUACE (YAROWSKI, 1995)

Jakmile množiny
zařazených slov přestanou
narůstat, zastav.





ONE SENSE PER DISCOURSE

- ... if a polysemous word such as *sentence* appears two or more times in a well-written discourse, it is extremely likely that they will all share the same sense.

LITERATURA

- Wikipedia contributors. **Yarowsky algorithm**. Wikipedia, The Free Encyclopedia. August 20, 2020, 19:47 UTC. Available at: https://en.wikipedia.org/w/index.php?title=Yarowsky_algorithm&oldid=974046729. Accessed November 2, 2020.
- William A. Gale, Kenneth W. Church, and David Yarowsky. **One sense per discourse**. In Proceedings of the workshop on Speech and Natural Language (HLT '91). Association for Computational Linguistics, USA, 233–237. 1992. DOI: <https://doi.org/10.3115/1075527.1075579>