

**Je muž s 204 centimetry vysoký
nebo malý?**

Kateřina Nová

Článek a autoři

- ▶ **Is a 204 cm Man Tall or Small ? Acquisition of Numerical Common Sense from the Web**
- ▶ Katsuma Narisawa, Yotaro Watanabe, Junta Mizuno, Naoaki Okazaki and Kentaro Inui
- ▶ The 51st Annual Meeting of the Association for Computational Linguistics (ACL 2013)
- ▶ Sofia, Bulgaria, August 4-9, 2013
- ▶ <http://aclweb.org/anthology//P/P13/P13-1038.pdf>



Numerický common sense

- ▶ Jak poznat, zda je číslo v nějakém kontextu velké, malé či normální?
- ▶ Doposud se tomu nevěnovala příliš velká pozornost, výzkumy se zabývaly spíše samotnými číselnými hodnotami než jejich významem v určitém kontextu
 - ▶ např. najít v textu odpověď na otázku „Kolik váží Airbus A380?“



Příklad

- ▶ t: Before long, 3b people will face a water shortage in the world.
- ▶ h: Before long, a serious water shortage will occur in the world
- ▶ **3b people face a water shortage.**
 - ▶ = 3,000,000,000 people face a water shortage.
 - ▶ => many people face a water shortage
 - ▶ => a serious water shortage



Jak na to?

- ▶ Získat data z webu: číselná vyjádření a jejich kontext
- ▶ Kontext = sloveso a jeho argumenty
 - ▶ např. věta „3b people face a water shortage.“



Jak na to?

- ▶ Konverze číselných vyjádření na sémantickou reprezentaci
 - ▶ „Roughly three thousand kilograms of meats have been provided every day.“
 - ▶ Převod číselné hodnoty: *three thousand kilograms* -> 3 000 kg -> 3 000 000 g
 - ▶ Nastavení modifikátoru: *roughly* -> „about“

Numerical Expression	Semantic representation		
	Value	Unit	Mod.
<i>about seven grams</i>	7	g	about
<i>roughly 7 kg</i>	7000	g	about
<i>as heavy as 7 tons</i>	7×10^6	g	large
<i>as cheap as \$1</i>	1	\$	small
<i>30–40 people</i>	[30, 40]	nin (<i>people</i>)	
<i>more than 30 cars</i>	30	dai (<i>cars</i>)	over
<i>7 km per hour</i>	7000	m/h	

Table 2: Normalized representation examples

String	Operation
gram(s)	set-unit: 'g'
kilogram(s)	set-unit: 'g'; multiply-value: 1,000
kg	set-unit: 'g'; multiply-value: 1,000
ton(s)	set-unit: 'g'; multiply-value: 1,000,000
nin (<i>people</i>)	set-unit: 'nin' (<i>person</i>)
about	set-modifier: 'about'
as many as	set-modifier: 'large'
as little as	set-modifier: 'small'

Table 3: An example of unit/modifier dictionary

Jak na to?

- ▶ 2 přístupy:
 - ▶ Odhad rozložení čísel vyskytujících se v daném kontextu a na jejich základě zjištění hodnoty daného čísla
 - ▶ Tato možnost vyžaduje dostatečnou „databázi hodnot“, aby bylo možné odhadnout, zda je daná hodnota nízká, vysoká či normální.
 - ▶ Textové vzory, kde mluvčí/pisatel explicitně vyjadřuje svůj názor na číselnou hodnotu
 - ▶ Ve větě se vyskytují výrazy jako „only“, „as many as“, „as few as“ apod.
- ▶ Oba přístupy se ukazují jako efektivní



Výsledky

- ▶ Relativně vysoká přesnost odhadů
- ▶ Odhad rozložení čísel je dobrý na určení „normálních“ hodnot
- ▶ Textová vodítka jsou lepší v určování nízkých a vysokých hodnot



Výsledky

No.	System	Gold	Sentence	Remark
1	small	small	I think that <u>three men</u> can create such a great thing in the world.	Correct
2	normal	normal	I have <u>two cats</u> .	Correct
3	large	large	It's <u>above 32 centigrade</u> .	Correct
4	large	large	I earned <u>10 million yen</u> from horse racing.	Correct
5	small	normal	There are <u>2 reasons</u> .	<i>Difficulty in judging small.</i> Since a few people say, "There are only 2 reasons," our approach predicted a <i>small</i> label.
6	small	large	<u>Ten or more people</u> came, and my eight-mat room was packed.	<i>Difficulty in modeling the context</i> because this sentence omits the locational argument for the verb <i>came</i> . We should extract the context as <i>the number of people who came to my eight-mat room</i> instead of <i>the number of people who came</i> .
7	small	normal	I have <u>two friends</u> who have broken up with their boyfriends recently.	<i>Difficulty in modeling the context.</i> We should extract context as <i>the number of friends who have broken up with their boyfriends recently</i> instead of <i>the number of friends</i> .
8	small	large		<i>Lack of knowledge.</i> We extract the context as <i>the number of heads of a turtle</i> , but no corresponding information was found on the Web.

Table 6: Output example and error analysis. We present translations of the sentences, which were originally in Japanese.



Děkuji za pozornost.