Predictive Writing

What is smart on smart phones . . .

Zuzana Nevěřilová, Barbora Ulipová

Natural Language Processing Centre, Faculty of Informatics
Masaryk University, Brno, Czech Republic

xpopelk@fi.muni.cz, b.ulipova@gmail.com

RASLAN 2014, Karlova Studánka, 2014-12-05

Predictive writing

- very usual on mobile devices:
  [Hasegawa et al., 2012] reported "average number of errors on software keyboards 4.55%, and the average number of hardware was 1.36%".
- less frequent on desktop computers
- we can benefit from the server performance and large memory

Custom data

on mobile devices:

- address book
- users' input (e-mails, SMS, IM)

desktop application:

- user-oriented
- genre-oriented

(Who) does predictive writing help?

motor impaired persons, users with dysgraphia, foreigners learning the language, and touchscreen users

but also “normal” users since nobody wants to type

N-gram model

minimum n-gram model for English 20,000 n-grams

big corpus $\rightarrow$ large n-gram model

in our case (czTenTen12):

- for $n = \{1, \ldots, 12\}$
- 150 M n-grams

Custom-built model

allow the user to upload their own texts

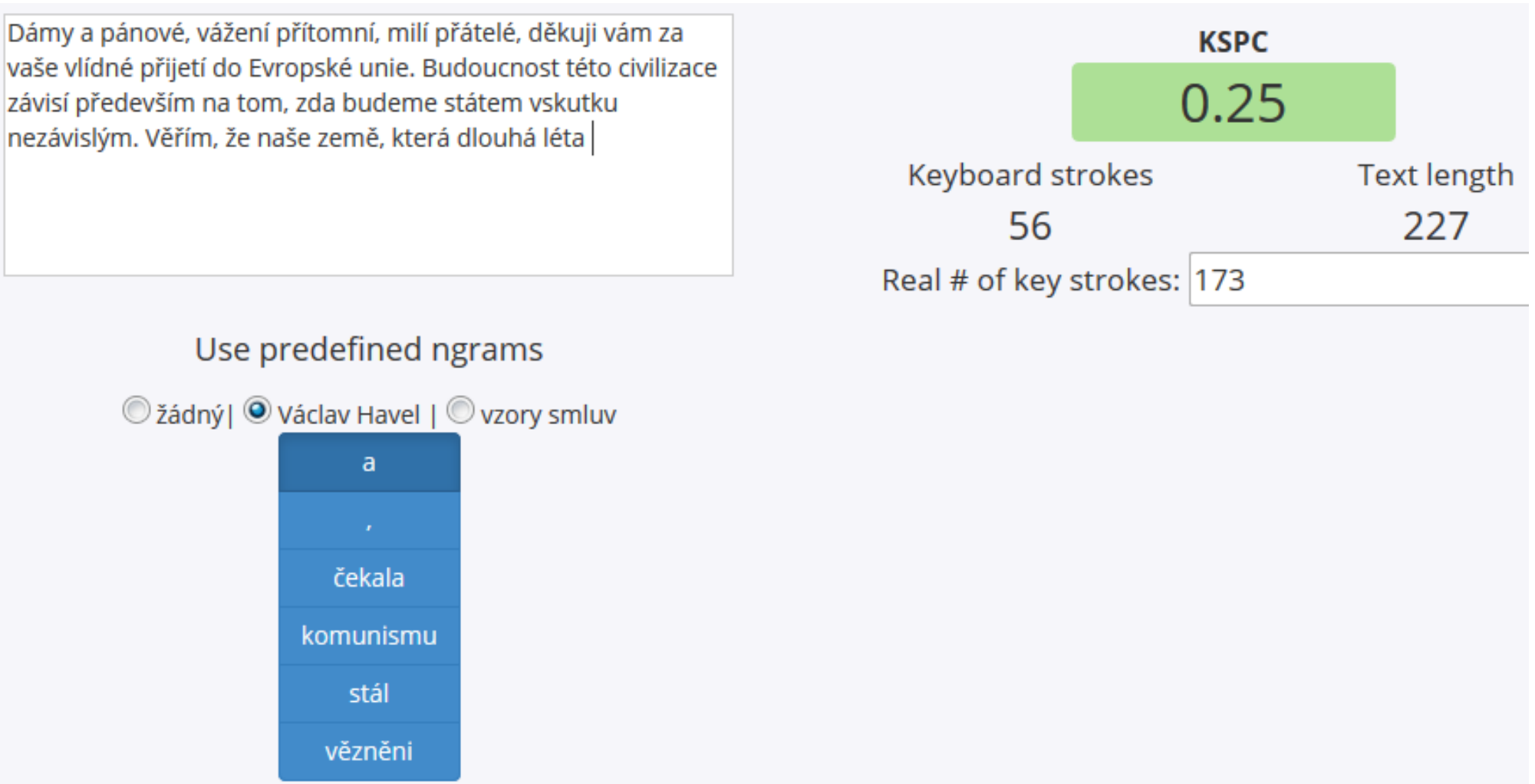
calculate n-grams $Score = FrequencyDistribution \cdot n^4$

custom-built model is suggested first

user data:

- particular user
- particular domain/genre

User interface



Server side

- each keystroke → server-side script
- the server-side performance matters
- searching in 150,000,000 n-grams using Daciuk finite state automata[1]

[1] http://galaxy.eti.pg.gda.pl/katedry/kiw/pracownicy/Jan.Daciuk/personal/fsa.html

Methods of measuring the results

- letter KSPC – number of keystrokes/text length
- real KSPC – keystrokes include control keys (arrows)

letter KSPC does not count control keys (arrows)
letter KSPC counts confirmation key (the Tab key)

letter KSPC describes the model
real KSPC describes the text input method

Other methods

- keystrokes/text length (KSPC without selection key)
- number of errors
- typing speed

Results

texts: Václav Havel speech, Václav Havel essay, contract specimen

Table : Resulting KSPC on four texts

text type	mortgage contract	contract of sale	H. speech	H. essay
letter KSPC	0.63	0.55	0.68	0.71
letter KSPC with custom-built model	0.62	0.54	0.67	0.70
real KSPC	0.77	0.64	0.81	0.84
real KSPC with custom-built model	0.70	0.62	0.81	0.77

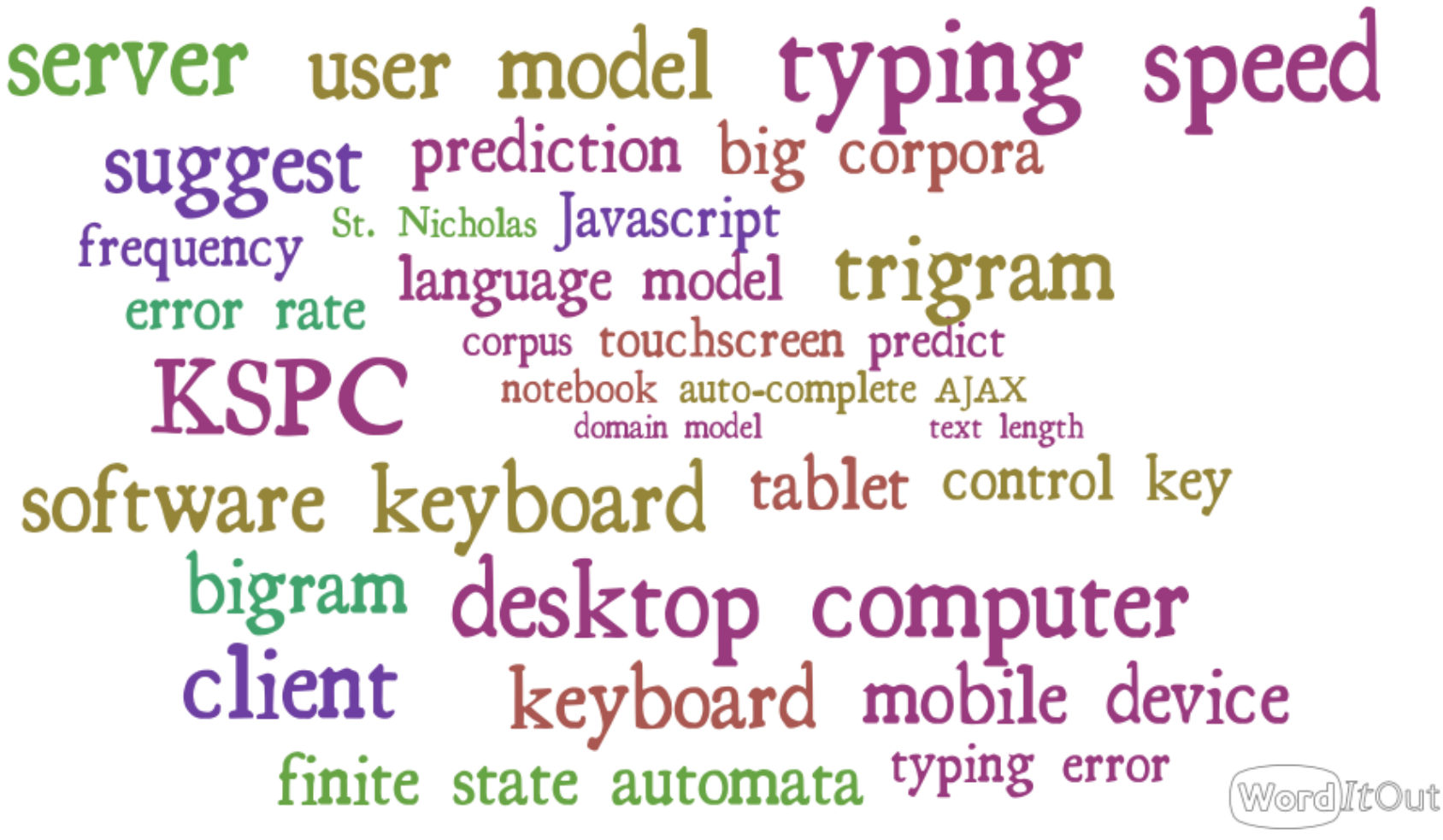
Comments on results

- average letter KPSC 0.64, real KSPC was 0.77 (comparable with WordTree [Badr and Raynal, 2009] which reports KSPC=0.71)
- custom-built n-grams: contracts more successful than Havel
- custom-built n-grams did not improve V.H. speech
- results dependent of writing technique

annotator	Bára	Zuzana
keystrokes/text length without custom-built model	0.51	0.46
keystrokes/text length with custom-built model	0.49	0.42
real KSPC without custom-built model	0.77	0.82
real KSPC with custom-built model	0.70	0.80

Future Work

- server side: negative n-grams (for spelling errors)
- client side: different text input method
- predict more tokens than one
- learning from user input


server
user model
typing speed
suggest
prediction
big corpora
St. Nicholas
Javascript
frequency
language model
trigram
error rate
corpus
touchscreen
predict
KSPC
notebook
auto-complete
AJAX
domain model
text length
software keyboard
tablet
control key
bigram
desktop computer
client
keyboard
mobile device
finite state automata
typing error
WordItOut


Badr, G. and Raynal, M. (2009).
WordTree: Results of a word prediction system presented thanks to a tree.
In Stephanidis, C., editor, *Universal Access in Human-Computer Interaction. Applications and Services*, volume 5616 of *Lecture Notes in Computer Science*, pages 463–471. Springer Berlin Heidelberg.


Hasegawa, A., Yamazumi, T., Hasegawa, S., and Miyao, M. (2012).
Evaluating the input of characters using software keyboards in a mobile learning environment: A comparison between software touchpanel devices and hardware keyboards.
In *IEEE International Conference on Wireless, Mobile, and Ubiquitous Technology in Education*, pages 214–217.