



# Proč není snadné přečíst nápis z fotky a proč bychom to měli chtít

Jindřich Libovický

Propaguj školu na své škole  
Gymnázium Kladno, 23.10. 2014

Co je rozpoznávání textu?

Dekódování textu

Počítačová lingvistika v rozpoznávání

Závěr

Co je rozpoznávání textu?

Dekódování textu

Počítačová lingvistika v rozpoznávání

Závěr

Co je rozpoznávání textu?

Dekódování textu

Počítačová lingvistika v rozpoznávání

Závěr

# Scene Text Recognition



Zdroj: ICDAR Robust Reading Dataset



(200, 532) (358, 575)

**FREEDOM**

# K čemu je to dobré?

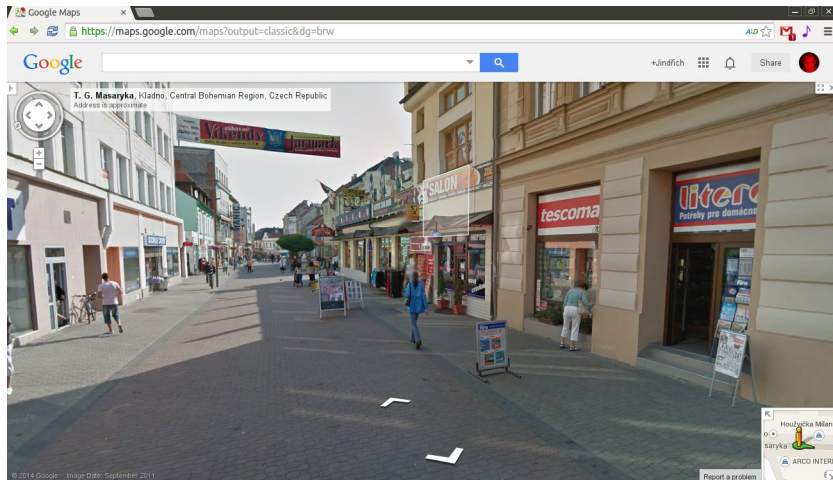


# K čemu je to dobré



# K čemu je to dobré

Co je rozpoznávání textu?





# Obsah

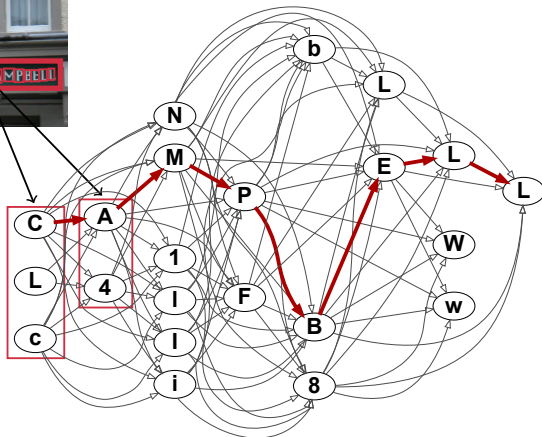
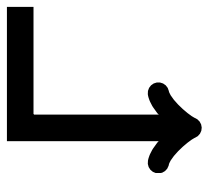
Co je rozpoznávání textu?

Dekódování textu

Počítačová lingvistika v rozpoznávání

Závěr

## Výsledek zpracování obrazu



# Hledání nejlepší cesty

Malostranské náměstí x

https://maps.google.com/maps?output=classic&dg=brw

Google

Get directions My places

A Malostranské náměstí 25, Prague, Czech Rep

B náměstí Edvarda Beneše 1573, Kladno, Czech

GET DIRECTIONS

Suggested routes

Route	Distance	Time	Traffic
R6	32.4 km	30 mins	In current traffic: 31 mins
R7 and Route 61	28.6 km	34 mins	No traffic information

Driving directions to náměstí Edvarda Beneše 1573, 272 01 Kladno, Czech Republic

This route has tolls.

A Malostranské náměstí 2/25  
118 00 Praha, Czech Republic

- Head west on Malostranské náměstí
- Turn left to stay on Malostranské náměstí
- Turn right onto Karmelitiská

Map data ©2014 Google - Edit in Google Map Maker Report a problem

# Jak to měřit?

- ▶ na mapě vzdálenost, čas, ...

... ale co v grafu z písmenek

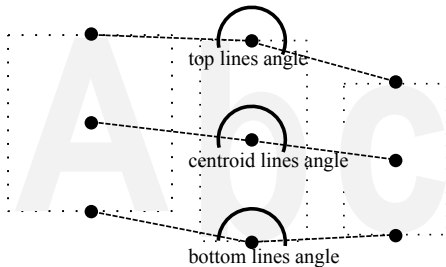
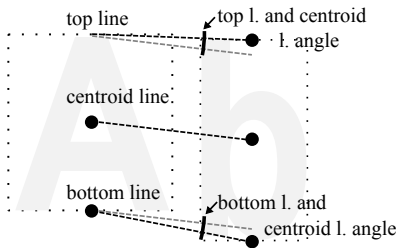
Co je rozpoznávání textu?

Dekódování textu

Počítačová lingvistika v rozpoznávání

Závěr

# Typografický model



# Model jazyka

- ▶ pravděpodobnost skupin písmen v jazyce  
 $P(\text{„řý“} \mid \text{čeština}) < P(\text{„ří“} \mid \text{čeština})$
- ▶ slovo je čitelné  $\Leftrightarrow$  lze rozdělit na slabiky  
ling-vis-ti-ka
- ▶ když slovo existuje, bude ve slovníku nebo na webu

# Výsledky rozpoznávání





# Shrnutí

- ▶ rozpoznávání textu na fotografiích není snadné, okolo 50 %
- ▶ pořádají se soutěže
- ▶ znalost jazyka dokáže výrazně pomoci

Děkuji za pozornost.