

# Analýza zložitosti textov

Martin Štrbák, Jakub Mrocek

# Obsah prezentácie

- Cieľ experimentov a prác
- Zložitosť textu, meranie, metriky
- Zhodnotenie doterajších experimentov
- Aplikácia poznatkov a návrh finálneho experimentu

# Ciele prác

- Vedieť určiť zložitosť textu ako celku pre čitateľa
  - Hypotéza: *“Pomocou EEG vieme rozlíšiť aktivitu čítania od stavu pokoja.”*
  - Hypotéza: *“Kognitívna záťaž v bezchybnom texte bude nižšia ako po zavedení chýb.”*
  -
- Identifikovať problémové časti z pohľadu zrozumiteľnosti
  - Hypotéza: *“Spôsob pohybu pohľadu čitateľa, ktorý narazí v texte na chybu sa bude líšiť od pohľadu v bezchybnom texte a zmeny budú môcť byť charakterizované zmenou hodnôt merateľných veličín o pohľade čitateľa.”*
  - Hypotéza: *“Oko čitateľa bude rozdielne reagovať na rôzne druhy chýb zavedených do čítaného textu.”*

# Zložitosť textu

- Subjektívna
- Závislá od štýlu, kompozície viet, použitých výrazov...
- Vieme vypočítať metriky na určenie predpokladanej zložitosti
- Sledovaním používateľa počas čítania chceme overiť subjektívnu zložitosť

# Použité metriky

- “Štandardné” metriky využívajúce štatistické hodnoty textu (počet písmen, slov a viet)
  - Coleman–Liau
  - Automated readability index
- Využite frekvencie z národného korpusu
  - Vypočítané skóre pre každé slovo. Čím nižšia frekvencia, tým vyššie skóre.
- Výsledky porovnáme s inovovanou mierou Nestlerovej – Průchu – Pluskala
  - Používaná na určovanie zložitosti SVK textu, ťažko sa počíta

# Predchádzajúci experiment

- Zavádzali sme lexikálne chyby (zámena písmen na konci / začiatku slova)
- 2 spôsoby dosiahnutia zložitého textu:
  - sémantický
    - naozaj náročné texty
  - typografický
    - zavedením chýb
- Priebežná kontrola pozornosti pomocou otázky na obsah textu  
=> čítanie s porozumením
- Po každom texte na 10 sekúnd čierna obrazovka

# Poznatky z prvého experimentu

- Zavedené chyby boli príliš umelé, text sa vďaka nim nestal náročnejší, akurát nezmyselný
- Tie isté texty sme nemali prečítané s chybami aj bez chýb, nevedeli sme priamo porovnať metriky
- Texty na konci experimentu mohli byť prečítané inak kvôli únave
- Bolesť hlavy pri dlhom čítaní (EEG maska tlačí)

# Priebeh experimentu

- Ľahké a ťažké (sémanticky)
  - z každej kategórie 10 textov
  - približne 30 minút čítania
- Náhodné poradie textov
- Rozdelenie textovej množiny na bez/chybné
  - chyby aj v ľahkých textoch
- Medzi čítaním 10 sekúnd pozeranie na krížik v strede obrazovky



# Typy chýb, ktoré do textu zavedieme

- Záměna slova za synonymum z druhej strany frekvenčného spektra
  - Básnik používa inakšie slová
  - Básnik používa onakvejšie slová
- Záměna slova, ktorá spôsobí logický odpor vo vete
  - “Lod’ sa kývala v rozbúrenom mori.”
  - “Lod’ sa smiala v rozbúrenom mori.”

# Výzva do diskusie

- Zavádzanie chýb, typy, frekvencia, rôznorodosť chýb v rámci textu (iba jeden typ chyby alebo oba)?
- Stav neaktivity, ponechať iba krížik alebo striedať obrázky / aktivity?
- Overenie porozumenia textu - potrebujeme ho vôbec? Ako ho využiť, zohľadniť?