Multi-Level Feature Extraction for Spelling Correction

Johannes Schaback (Technische Universität Berlin), Fang Li (Shanghai Jiao Tong University)

Our Spell Checker framework at a glance:

▶ plain machine learning
▶ non-word errors and real-word errors
▶ no “confusion lists”

Candidate discrimination via evidence from

▶ Character Level
▶ Phonetic Level
▶ Word Level
▶ Syntactic Level
▶ Semantic Level

Final discrimination by Support-Vector Machine

→ 3% improvement over competitive spell checkers