

Korrekturprogramme

—“Nicht-Wörter“

Dozentin: Wiebke Petersen

Referentin: Yun Ji

Arten der Korrekturprogramme

- Programme zur Korrektur von “Nicht-Wörtern“
- Programme zur kontextabhängigen Korrektur
- Grammatikkorrekturprogramme

Definition von „Nicht-Wörtern“ (*non-words*):

Zeichenketten, die in der Sprache nicht vorkommen.

- Auslassung (z.B. Compute)
- Ersetzung (z.B. Komputer)
- Einfügung (z.B. Chomputer)
- Vertauschung/Transposition (z.B. Comptuer)

Erkennung der Wörter

Lexikon (*dictionary*):

Ein Inventar von Wörtern,
Wortbestandteilen oder Wortgruppen, zu
denen verarbeitungsrelevante
Informationen gegeben werden.

Probleme bei bloßer Verwendung der Lexika

- hochfleckierte & agglutinierende Sprachen
(Türkisch, Ungarisch)
- Komposition & Derivation
z.B. Korrektur-programm
Richtig-keit

Korrigierung (1): Suche nach anpassenden Korrekturvorschlägen

- Ein falsch eingefügter Buchstabe:
N Operationen
- Eine Transposition: N-1 Operationen
- Ein ausgelassener Buchstabe: $k(N+1)$ Operationen
- Ein falsch ersetzter Buchstabe: $(k-1)N$ Operationen

*Voraussetzung: ein fehlerhaftes Wort mit Länge N
(k=Zahl der Buchstaben in einer Sprache)

Korrektur(2): Auswahl des optimalen Vorschlags

- Edit Costs: Costs zwischen der fehlerhaften Sequenz und dem Korrekturvorschlag

z.B. Edit Costs(ih,ich)=1

- **Minimal Edit Distance:** Das fehlerhafte Wort kann mit möglichst wenigen Änderungen in den Korrekturvorschlag überführt werden.

z.B. Fehlerhaftes Wort: * Libe

Korrekturvorschläge: ★Liebe (1 Auslassung)

Leib (2 Vertauschungen)

Menschliche Interaktion (James L. Peterson 1980):

- Ersetzung (Replace)
- Ersetzung und Erinnerung (Replace and Remember)
- Akzeptierung (Accept)
- Akzeptierung und Erinnerung (Accept and Remember)
- Edit (Edit)

Beispiel:Ispell & Aspell

- Ispell: Umlaut(Bauer als *Baür)
„ß“-Problem
- Aspell:
 - Vorteil:bessere Vorschläge (trouble für *trubble)
 - Nachteil: nur für Englisch

Problem: bei Sprachen mit Zeichenschrift

- Chinesisch:
 - Schreibfehler ≠ Tippfehler: kaum Nicht-Wörter und kaum „real word“ error

Beispiele:

步vs.*步 涼vs.*涼

戍shù 戍xū 戊wù 成chéng 戎róng 戒jiè

(*diese zwei Zeichen sind Nicht-Wörter im Chinesisch, aber richtige Zeichen im Japanisch.)

Vielen Danke!