

Pb

On ne cherche plus ici à aligner deux mots, mais des facteurs de ces mots.

Plus précisément cherche si un facteur de l'un ressemble beaucoup à un facteur de l'autre, il ne s'agit pas d'aligner tous les facteurs possibles.

Idée

La distance est une mauvaise mesure pour ce problème car elle ne mesure que les coûts, ne fait que pénaliser.

Ainsi si on cherche deux facteurs de distance minimale on a de forte chance de tomber sur un facteur réduit à une lettre. En effet aligner A et A coûte 0.

Mais ce n'est pas intéressant car on voudrait aussi que les facteurs soient longs. C'est pourquoi il faut valoriser les cas où on aligne vraiment, c-à-d les substitutions triviales.

SUB

On considère à nouveau cette opération symétrique. C'est cette opération qui doit faire apparaître la sélection des "caractères en phase".

$$\text{sub}(a, b) = \begin{matrix} \text{coût de substitution } a \text{ à } b \\ \text{ou vice-versa} \end{matrix} = \begin{cases} > 0 & \text{si } a \neq b \\ < 0 & \text{si } a = b \end{cases}$$

DEL

On considère que le "coût" de la suppression dépend de la lettre supprimée, l'améliora par rapport à ce qu'on faisait avant, pas tellement liée au fait

$$\text{del}(a) = \text{coût d'une suppression d'un } a \quad \underline{\leq 0}$$

qu'on passe en local, mais plus au fait qu'on n'a plus des contraintes pour que les soit une distance...

INS

Ici aussi on considère que le "coût" dépend de la lettre insérée

$$\text{ins}(a) = \text{coût de l'insertion d'un } a \quad \underline{\leq 0}$$

Def Le score d'un alignement (\tilde{u}, \tilde{v}) de u et v est défini comme

$$S(\tilde{u}, \tilde{v}) = \sum_{i=1}^{|\tilde{u}|} s(\tilde{u}_i, \tilde{v}_i) \quad \text{où } s(a, b) = \begin{cases} \underline{\text{del}}(a) & \text{si } b = - \\ \underline{\text{ins}}(b) & \text{si } a = - \\ \underline{\text{sub}}(a, b) & \text{sinon} \end{cases}$$

ex

$$\begin{aligned} \underline{\text{del}}(A) &= -1 & \underline{\text{ins}}(A) &= -2 \\ \underline{\text{del}}(B) &= -2 & \underline{\text{ins}}(B) &= -1 \\ \underline{\text{del}}(C) &= -2 & \underline{\text{ins}}(C) &= -1 \end{aligned}$$

sub

	A	B	C
A	3	-5	-5
B	-5	2	-1
C	-5	-1	2

$$\begin{aligned} \tilde{u} &= BBA_CBCAB_ _ C_ \\ \tilde{v} &= B_BACCB_BCCAB \\ & \quad \underline{2} \quad \underline{-2} \quad \underline{-5} \quad \underline{-2} \quad \underline{2} \quad \underline{-1} \quad \underline{-1} \quad \underline{2} \quad \underline{-1} \quad \underline{-1} \quad \underline{-5} \quad \underline{-1} \end{aligned}$$

$$S(\tilde{u}, \tilde{v}) = -1$$

Def La similarité entre deux mots u et v est alors le score maximal d'un alignement de u et v .

$$\underline{\text{Sim}}(u, v) = \max \{ S(\tilde{u}, \tilde{v}) \mid (\tilde{u}, \tilde{v}) \text{ alignement de } u \text{ et } v \}$$

Def On appelle meilleur alignement local de x et y , un alignement de facteurs de x et y resp. de score maximal.

ex

$$\begin{aligned} \underline{\text{del}} &\equiv -1 & \swarrow & \text{coût indépendants de la lettre supprimée ou insérée} \\ \underline{\text{ins}} &\equiv -1 & \searrow & \\ \underline{\text{sub}}(x, y) &\equiv -1 \text{ si } x \neq y & & \\ &\equiv +3 \text{ si } x = y & & \end{aligned}$$

$u = \text{ENCADREMENT}$
 $v = \text{ACCABLER}$

$$\begin{aligned} \tilde{u} &= \text{ENCADREMENT} \\ \tilde{v} &= \text{ACCABLER} _ _ _ \\ & \quad \underline{-1} \quad \underline{-1} \quad \underline{+3} \quad \underline{+3} \quad \underline{-1} \quad \underline{-1} \quad \underline{+3} \quad \underline{-1} \quad \underline{-1} \quad \underline{-1} \quad \underline{-1} \end{aligned}$$

le meilleur alignement local est $\begin{matrix} \text{CADRE} \\ \text{CABLE} \end{matrix}$
de score $3+3-1-1+3 = \underline{7}$