

Algorithme projection simple

Si 2 Terms correspondent à un même synset \Rightarrow (même N^o)
Ils sont synonymes.

Début

Pour (chaque Ph de d) faire

W mot à analyser}

Tantque (non fin (Ph)) faire

1. Projeter W sur les \neq domaines

2. déterminer l'ensemble $E_{\text{complet}W}$ contenant les concepts $C_j^{Di} \in$ aux \neq domaines commençant par le mot W ($j=1,N$)

3. Si $E_{\text{complet}W} \neq \emptyset$ alors

3.1 ordonner les C_j^{Di} par ordre décroissant de leur

taille $\{|C_1^{Di}| \geq |C_2^{Di}| \geq \dots \dots |C_N^{Di}|\}$, $E_{\text{complet}W} = C_j^{Di}$

3.2 trouver le concept (ou les concepts, s'il ya plusieurs concepts de même taille) le plus long (ce concept va correspondre à un terme T de la phrase).

3.2.1 si le concept trouvé existe dans plusieurs domaines alors l'affecter à tous ces domaines.

3.2.2 pour chaque domaine donné Di ,

3.2.2.1 Si il existe plusieurs concepts ayant la même longueur, alors affecter ces concepts à EW_{ambigus}^{Di}

Sinon

affecter ce concept à E^{Di}

finsi

finpour

Finsi

Passer er au mot suivant W à analyser (W qui vient après le terme T)

Fintque

Finpour

Procédure filtrage (désambiguisation)

Fin

Remarques pour la projection

T identique à C.

2

lors de la projection on doit garder pour chaque phrase et pour chaque paragraphe :

- 1/ les concepts extraits,
- 2/ leur ordre (ordre des termes correspondants à ces concepts).

Procédure Filtrage { Dans chaque domaine, parmi les concepts candidats pour w, associer un seul concept à W }

Entrée :

$E_{ambigus}^{Di} = \{c_1, c_2, \dots\}$ concepts candidats pour W

ph (phrase courante où w apparait.)

pg (paragraphe courant où w apparait.)

Sortie

C (concept retenu pour w)

Début

Pour chaque domaine D^i faire

Pour tout ensemble $EW_{ambigus}^{Di}$ faire

c_1, c_2, \dots, c_n (concepts candidats associés à w obtenu par l'algorithme projection)

1. Trouver v_g (le voisin non ambigu de gauche, le plus proche de w dans le document)

2. Trouver v_d (le voisin non ambigu de droite, le plus proche de w dans le document)

3. Extraire cv_d (le concept associé au voisin de droite de w)

4. Extraire cv_g (le concept associé au voisin de gauche de w)

5. Calculer $Min-dist((c_i, c_{v_g}), (c_i, c_{v_d}))$

Finpour

affecter le concept retenu à E^{Di}

finpour

Fin.

Remarques

- Les voisins, on les cherche d'abord dans la phrase, puis dans le paragraphe.

- Si dans une phrase W ne contient qu'un seul voisin (celui de gauche ou celui de droite) alors ne considérer que ce voisin.

Algorithme filtrage

T terme à désambiguer

Pour chaque phrase ph faire

T ← premier terme ambigu de ph

Procédure filtrage (T)

debut

Tant que (non fin de ph) faire

Chercher $v_g(T)$

chercher $v_d(T)$

Si ($v_g(T)$ existe) et ($v_d(T)$ existe) alors

Si ($v_g(T)$ non ambigu) et ($v_d(T)$ non ambigu) alors

Calculer $C_{retenu} = \text{Min} - \text{dist}((c_i^T, c_{v_g}), (c_i^T, c_{v_d}))$

Procédure stock

Sinon

Si ($v_g(T)$ ambigu) et ($v_d(T)$ ambigu) alors

Filtrage($v_d(T)$)

Filtrage(T)

Sinon

Si ($v_g(T)$ ambigu) alors

Considérer uniquement $v_d(T)$

Procédure stock

Sinon

Considérer uniquement $v_g(T)$

Procédure stock

finsi

finsi

finsi

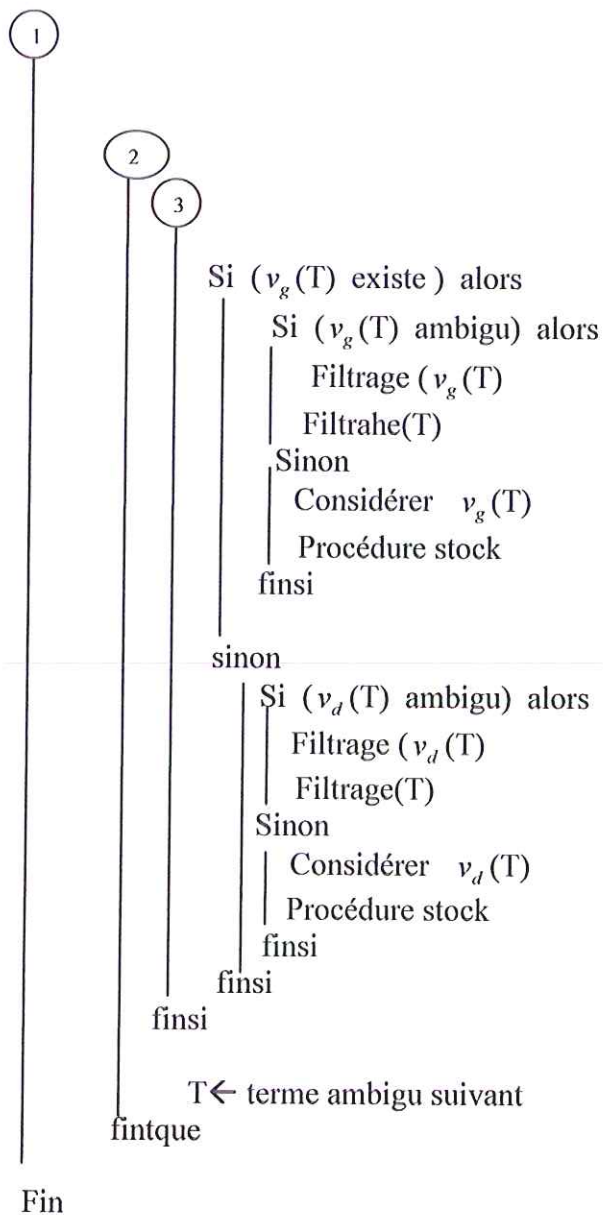
sinon

1

2

3

4



Procedure stock (C_{retenu})

Debut

$$EW_{\text{ambigus}}^{Di} = EW_{\text{ambigus}}^{Di} - C_{\text{retenu}}$$

$$E^{Di} = E^{Di} \cup C_{\text{retenu}}$$

fin

Remarques

1/ filtrage phrase par phrase pour chaque paragraphe. La limite est la phrase

2/ s'il reste des termes ambigus, refaire le même travail en considérant la limite, le paragraphe.

3/ s'il reste des termes ambigus, refaire le même travail et la limite sera le document.