

## **Quelques mots pour la préparation d'annotation des résultats de documents**

Yingsi LIANG

Dans le but d'annoter automatiquement les résultats présentés dans les articles en utilisant le logiciel Unitex, nous avons essayé de réfléchir les problèmes que nous allons rencontrer et trouvé les indices qui nous aident à annoter les résultats.

Pour avoir une idée générale des contenus à annoter, j'ai consulté attentivement deux documents de Denis Maurel, ils sont « Les transducteurs à sorties variables » et « Reconnaissance d'entités nommées : enrichissement d'un système à base de connaissances à partir de techniques de fouille de texte ».

Après avoir essayé d'annoter ces documents, j'ai remarqué qu'il y a plusieurs questions que devons envisager avant la construction des graphes de détection :

Premièrement, la définition plus précise pour « résultat », quel type de « résultat » que nous voulons annoter. Tout d'abord, je me concentre seulement sur les chiffres comme le score des expériences. Je trouve que c'est l'un des résultats le plus saillant et évident l'homme et la machine. Par contre, pour les comparaisons existantes dans les rapports, elles sont plus tôt apparues sous la forme narrative, c'est-à-dire comme une phrase complète. Dans ce cas-là, nous avons beaucoup plus de difficulté à décider le seuil d'annotation.

Deuxièmement, nous devons déterminer la limite ou la sphère de résultat que nous voulons. En un autre mot, nous réfléchirons si nous prenons en compte que les résultats que nous obtenons dans l'expérience du rapport, ou nous serons annoter tous les résultats présentés dans les articles même si c'est un résultat obtenu pour les autres expériences (plutôt apparaissent dans la partie l'état de l'art)?

Troisièmement, il est aussi nécessaire de éclairer le seuil de l'annotation. Prenons le cas des scores, nous allons seulement les chiffres ? C'est une solution simple, mais le problème est que ces chiffres n'ont pas de sens tout seule, du coup ces annotations ne permettraient pas de résumer les résultats ultérieurement. Nous pouvons peut-être annoter aussi les mots après les chiffres ? Ces mots peuvent nous aider à comprendre les chiffres, mais si c'est suffisamment complet ou pas ? Une autre solution nous pouvons tenter est annoter en même temps le sujet de résultat, les chiffres et les mots après les

chiffres, il me semble que les informations sont plus complètes. Cependant, la difficulté d'annotation va beaucoup augmenter ainsi que l'indication des rapports entre les unités sera compliquée.

Avec ces problèmes à discuter, j'ai annoté les deux articles avec la convention d'annotation ci-dessous :

Jeune	Résultat
Verte	Sujet de résultat
Rose	Indice d'annotation
Bleu	Les mots qui doivent prendre en compte, qui change le sens de résultat

À l'aide de ces annotations, j'ai imaginé plusieurs types de graphes à annoter les résultats :

1. Pour les chiffres, nous pouvons les annoter les chiffres qui se trouvent dans le paragraphe de « Résultat » (certain nombre de mots après le résultat).
2. Les verbes qui sont comme les indices d'introduire les résultats, tels que posséder, représenter, etc.
3. Les mots-clés qui peuvent présenter la comparaison, c'est-à-dire les comparatifs et les superlatifs.
4. Au début de la phrase, la structure « pour + ... + , » qui signifie le sujet d'analyse.
5. « verbe + ... + sur » sert à annoncer le sujet
6. Aussi les mots comme « respectivement » qui nous indique une structure parallèle entre des sujets et des résultats.

Finalement, pour la mesure d'évaluation, par instant je n'ai pas encore trop idée. Je pense que nous pouvons d'abord essayer la façon la plus générale la Précision et le Rappel ; nous ajoutons après les autres paramètres plus précis quand nous remarquons la nécessité de les prendre en compte.