

a French Standard-Based Resource for Keyphrase Extraction Evaluation

The TermITH - Eval experiment

This paper presents a dataset of evaluation scores assigned to automatically extracted keyphrases by human evaluators.

Along with the reference keyphrases, the manual evaluations can be used to validate new evaluation measures. Indeed, an evaluation measure that is highly correlated to the manual evaluation is appropriate for the evaluation of automatic keyphrase extraction methods

The TermITH - Eval dataset

400 bibliographical records from the Pascal and Francis databases of the French Institute for Scientific and Technical Information (INIST – CNRS)

-> 100 in each of the 4 specific domains: Linguistics / Information science / Archaeology / Chemistry

<teiHeader>

La cause linguistique </title>

L'objectif est de fournir une définition de base du concept linguistique de la cause en observant son expression. Dans un premier temps, l'A. se demande si un tel concept existe en langue. Puis il part des formes de son expression principale et directe (les verbes et les conjonctions de cause) pour caractériser linguistiquement ce qui fonde une telle notion. </abstract>

</keywords> </textClass>

<standOff>

<teiHeader>

<termEntry xml:id="plmilkwl>cause</term>
<termEntry xml:id="plmilkw2>définition</term>
<termEntryxml:id="plmilkw3>expression</term>

</listAnnotation>

<standOff>

</teiHeader>

<listAnnotation type="pertinence">

 <fs><f name="pertinence"><numeric value="2"/></f></fs>

 <fs><f name="pertinence"><numeric value="2"/></f></fs>

</listAnnotation>

<ns:listAnnotation type="silence">

 <fs><f name="silence"><numeric value="2"/></f></fs>

<note>implicite</note>

< fs > < f name = "silence" > < numeric value = "2" / > < / fs >

 </listAnnotation>

Bibliographical from the INIST database

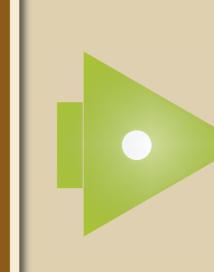
from which are taken into account:

- Title
- Abstract
- Reference keyphrases assigned by professionnal indexers

(on average: 10 keyphrases in linguistics and Information Sciences / 15 keyphrases in archaeology and chemistry)

Results of the term extraction process

TF-IDF Kea TopicRank



Manual evaluation of automatically extracted keyphrases regarding:

appropriateness silence

Score		Linguistics			Information Science			Archaeology			Chemistry		
Score	TF-IDF	KEA	TopicRank	TF-IDF	KEA	TopicRank	TF-IDF	KEA	TopicRank	TF-IDF	KEA	TopicRank	
2	35.3	37.2	37.1	34.7	34.2	36.3	46.0	49.9	51.6	50.9	54.0	53.7	
1 – non redundant	4.2	9.8	5.7	15.3	18.3	18.5	14.1	16.3	15.4	25.9	24.1	29.1	
1 – redundant	6.8	8.9	0.9	8.1	7.6	2.8	4.0	5.7	0.8	4.6	5.7	1.2	
0	53.8	44.0	56.3	41.9	39.9	42.4	35.9	28.1	32.2	18.7	16.3	16.0	

Table 1 Appropriateness ratios of TF-IDF, KEA and TopicRank on each specific domain

					_			_				
Score	Linguistics			Information Science			Archaeology			Chemistry		
Score	TF-IDF	KEA	TopicRank	TF-IDF	KEA	TopicRank	TF-IDF	KEA	TopicRank	TF-IDF	KEA	TopicRank
2	20.1	16.2	16.8	25.5	22.0	21.6	38.2	33.0	32.8	22.0	17.1	19.2
1	48.5	45.3	48.3	25.8	25.8	25.3	23.3	23.2	22.7	32.0	32.2	32.2
0	31.4	38.5	35.0	48.7	52.2	53.1	38.5	43.9	44.5	46.0	50.7	48.6

Table 2 Silence ratios of TF-IDF, KEA and TopicRank on each specific domain

Adrien Bougouin¹, Sabine Barreaux², Laurent Romary³, Florian Boudin¹, Béatrice Daille¹ ¹ Université de Nantes, LINA, Nantes, France, ² INIST-CNRS, Vandœuvre-lès-Nancy, France, ³ INRIA, Alpage project, Berlin, Germany E-mail: adrien.bougouin@univ-nantes.fr, sabine.barreaux@inist.fr, laurent.romary@inria.fr, florian.boudin@univ-nantes.fr, beatrice.daille@univ-nantes.fr



Format

annotations to class

together groups of

(experiment with

the <standOff>

discussion within

the TEI consortium)

entries compliant

standard 30042

(TBX, TermBase eXchange)

element under

Terminological

to ISO

TEI guidelines

annotations

Stand-off







