

# Résumé automatique d'articles scientifiques basé sur les citations

---

par Bruno Malenfant  
Directeur de thèse : Guy Lapalme

Université de Montréal

# Déroulement de la présentation

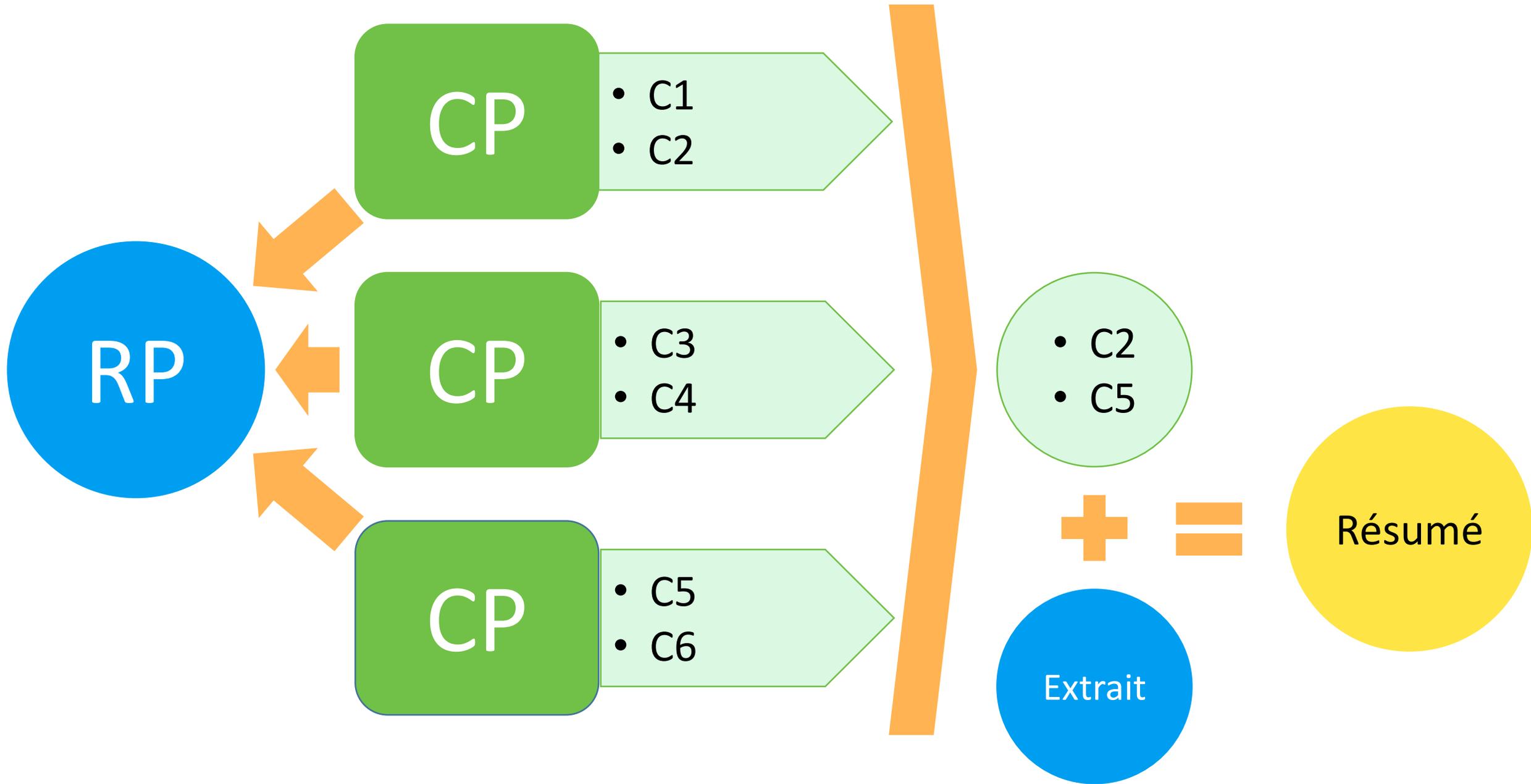
---

- || Présentation du problème
- || Revue de littérature.
- || Exemple.
- || Expérimentation.
- || Conclusion.

# Problème

---

- Un article contient déjà un résumé signalétique (abstract).
- Nous avons plusieurs techniques pour construire des résumés.
- Nous voulons ajouter à cette information, pouvoir montrer l'impact que l'article a eu.
- Cette information se situe dans les citations.



# Citance

---

Ensemble des phrases entourant une citation (Preslav et al.).

# Trouver les articles

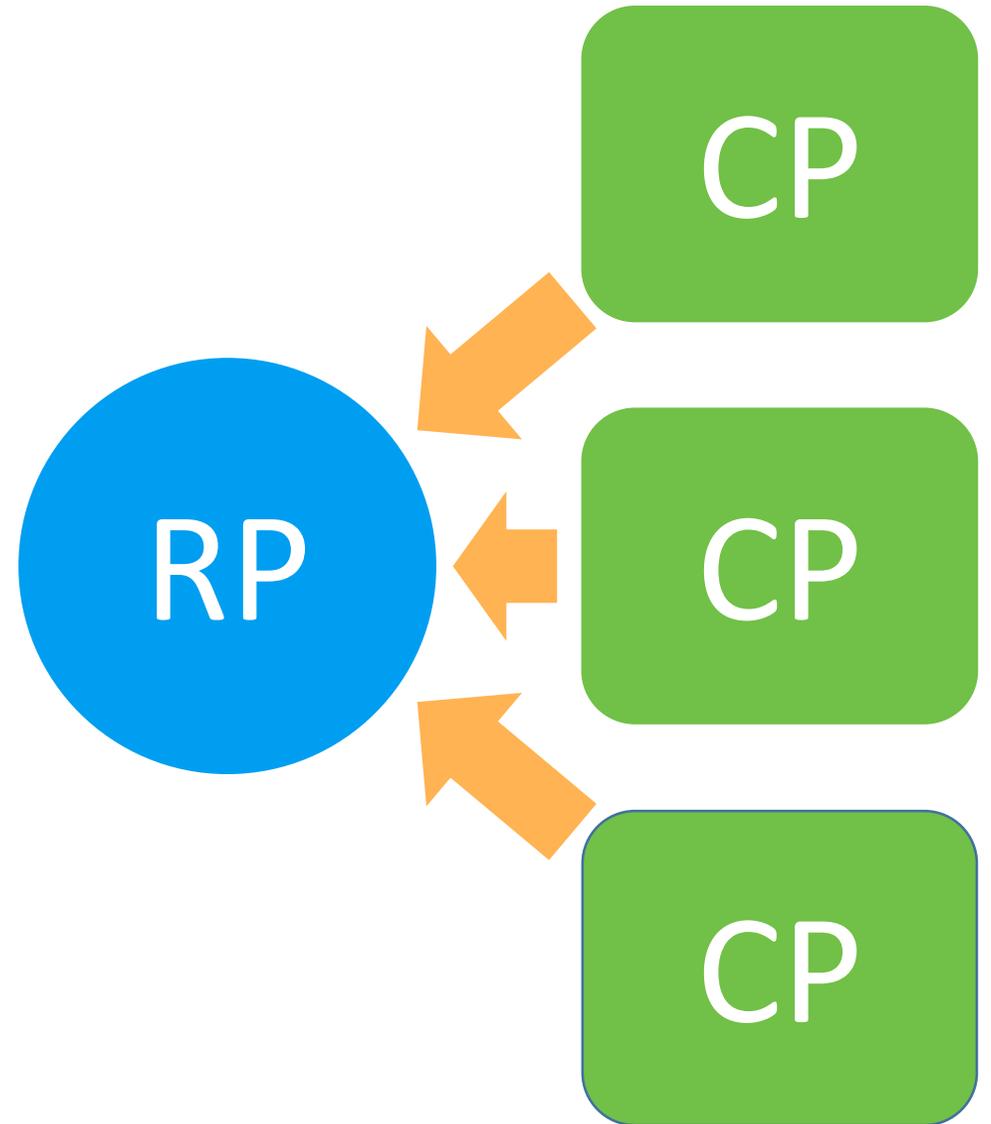
---

## 1. Outils

- a. Microsoft Academic Search
- b. CiteSeerX
- c. Google Scholar

## 2. Corpus

- a. AAN (18 000 articles)
- b. ESWC 14 (400 articles)
- c. TAC 2014 (550 articles)



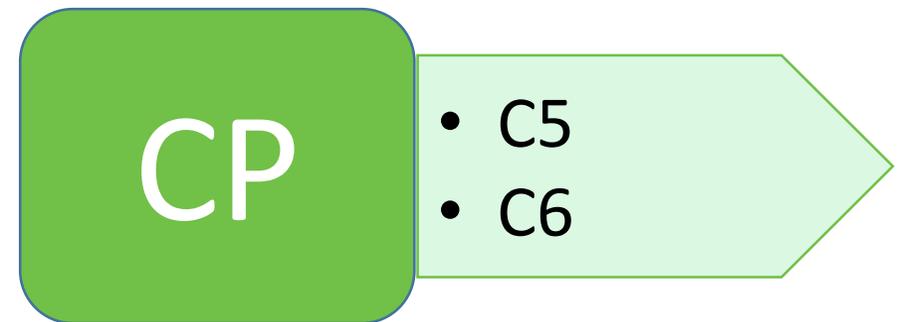
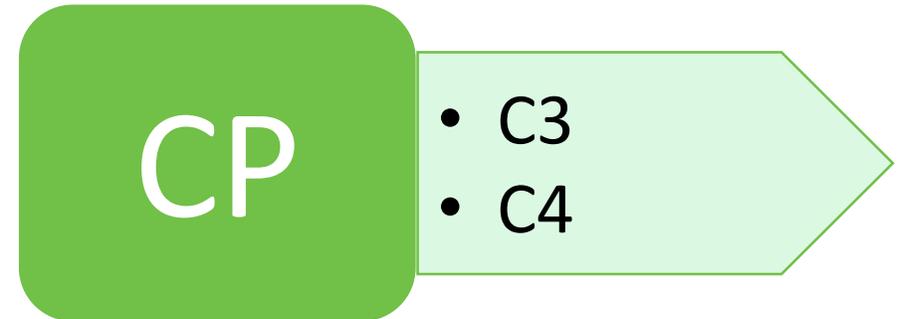
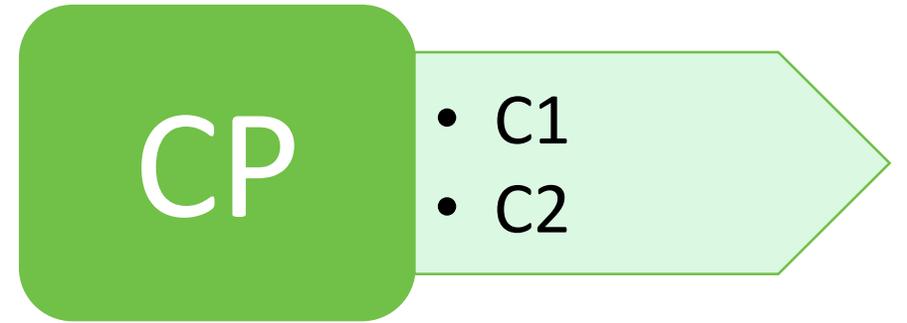
# Extraction des citances

---

1. Trouver les marqueurs.
2. Identifier les citances.

- Ou

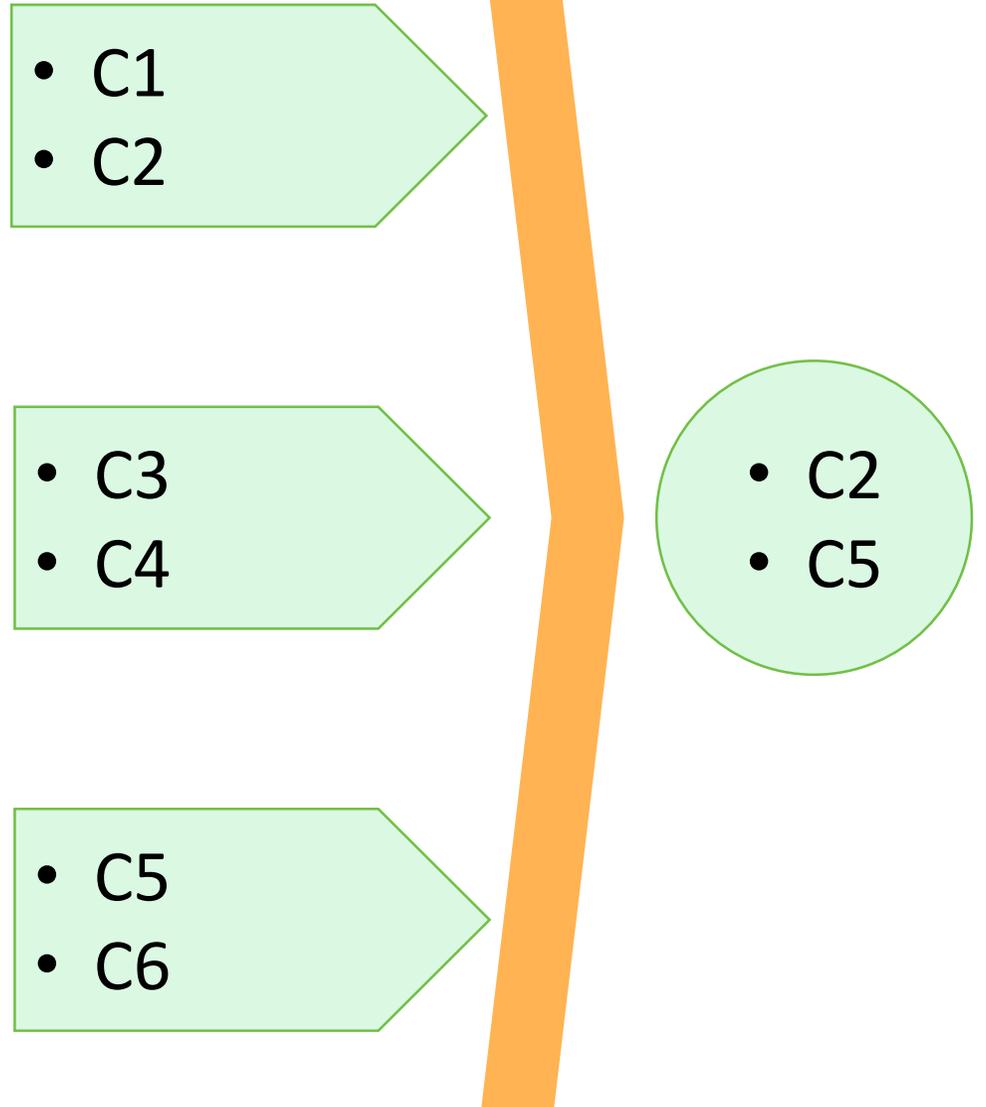
1. Texte déjà balisé (ESWC 14)



# Sélection des citances

---

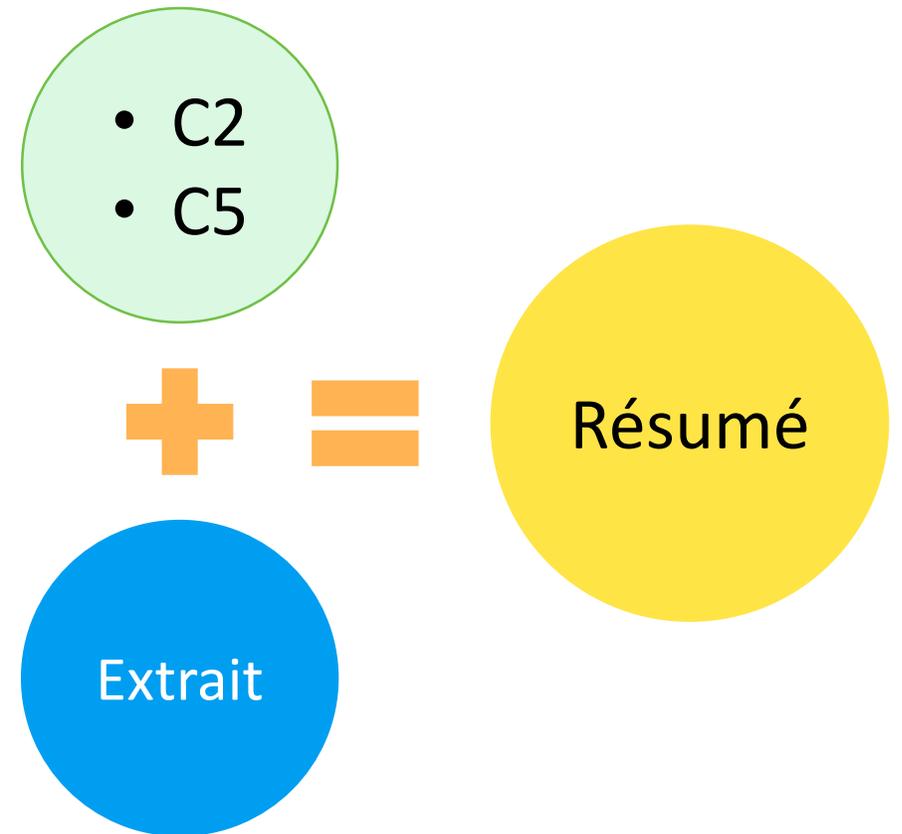
1. Identifier la facette des citances :
  - a. Base pour la recherche.
  - b. Indiquer le travail des autres.
  - c. Contraste avec une autre recherche.
2. Choisir les citances qui contiennent de l'information différente.



# Amélioration du résumé

---

1. Choisir des phrases apportant de l'information complémentaire.
2. Trouver les phrases référées.



# Facettes d'un article scientifique

---

- Simone Teufel
  - **aim, textual, own, background, other, contrast, basis.**
- Text Analysis Conference
  - **hypothesis, methods, results, implication, discussion.**
- CL-Summ
  - **aims, methods, results, implication.**

# Extraction des citations.

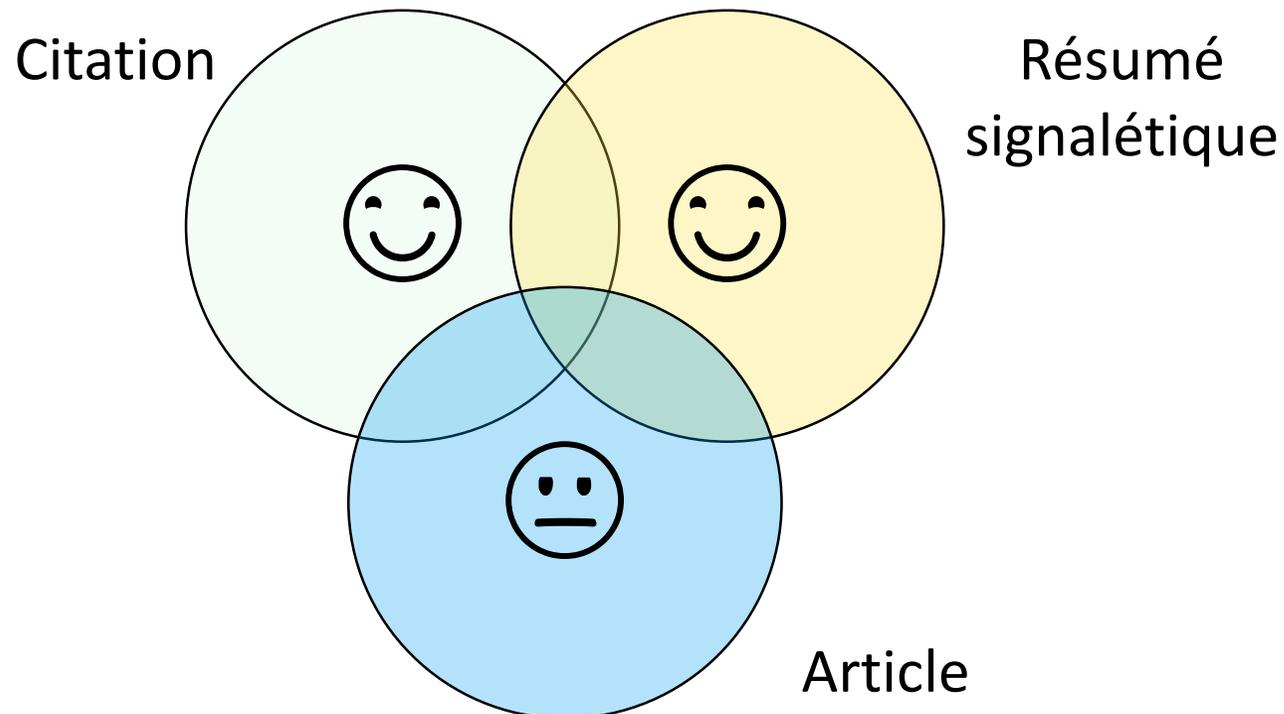
---

- Dominique Besagni et al. : patrons de reconnaissance.
- Vahed Qazvinian et Dragomir R. Radev : champs aléatoires de Markov.
- Brett Powley et Robert Dale : extraction alternée des citations et références.

# Résumé d'article multiples.

---

- Vahed Qazvinian et al. : utilisation des citations et de résumé signalétique.



# Exemple (Manuel)

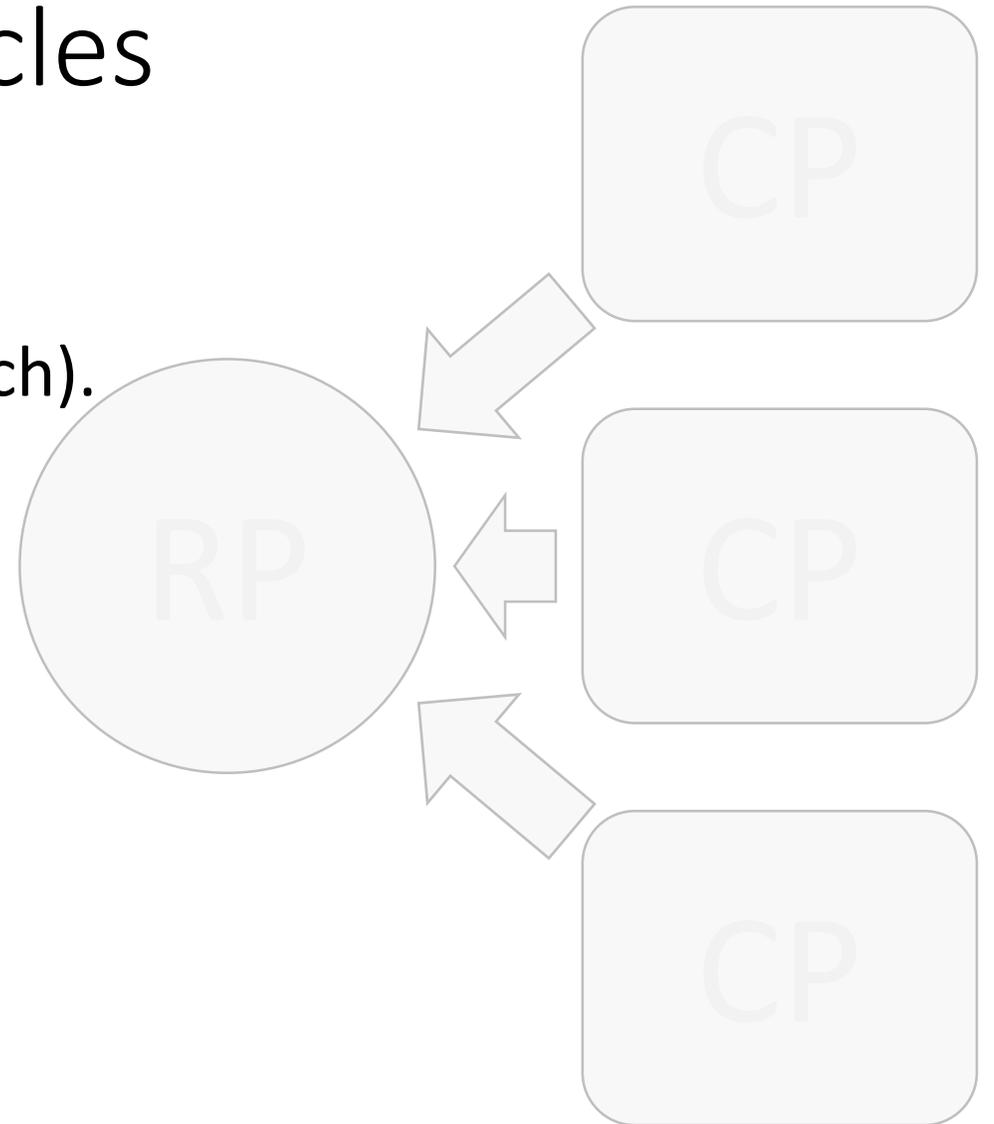
---

Monadic encapsulation of effects: a revised approach (extended version), Eugenio Moggi et Amr Sabry, 2001

# Exemple : Trouver les articles

---

- 26 citations (Microsoft Academic Search).
- 6 doublons.
- Sur 17 articles.



# Exemple : facettes des citations

---

## Other

*Moggi and Sabry [\*] prove syntactic type soundness for encapsulated lazy state.*

## Basis

*As described in the previous section, the design of FRGN takes inspiration from the work on monadic state[\*].*

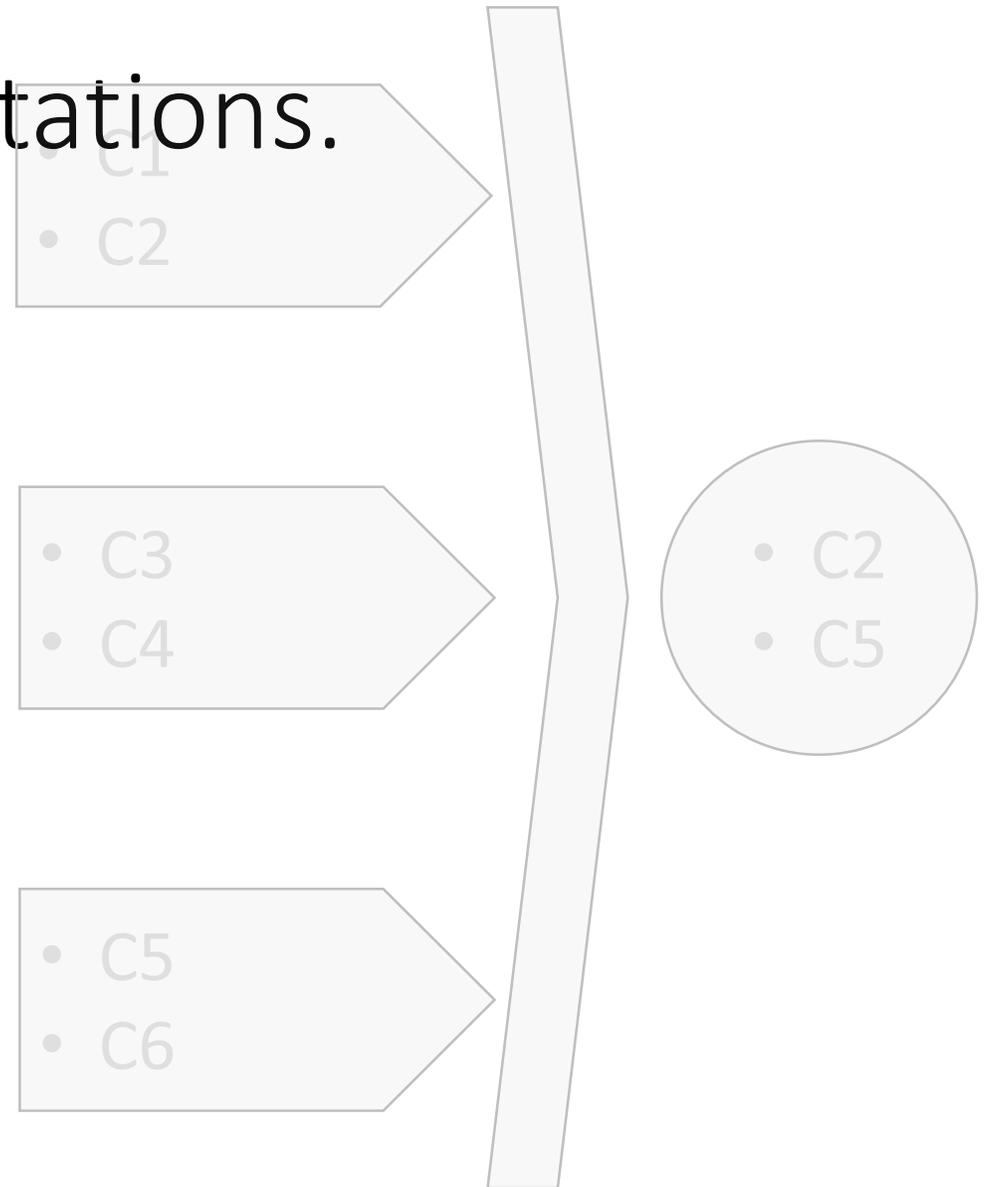
## Contrast

*While recent work [\*] has considered more general combinations of effects and monads, no work has examined the combination of regions and monads.*



# Exemple : sélection des citations.

- Besides supporting generic results that can be instantiated to particular effects at little or no cost, monads allow for a clear delineation of the scope of effects [\*].
- Secondly, side effects are cleanly encapsulated and equipped with a clearly delineated scope[\*].



# Exemple : résumé pour la facette **Other**

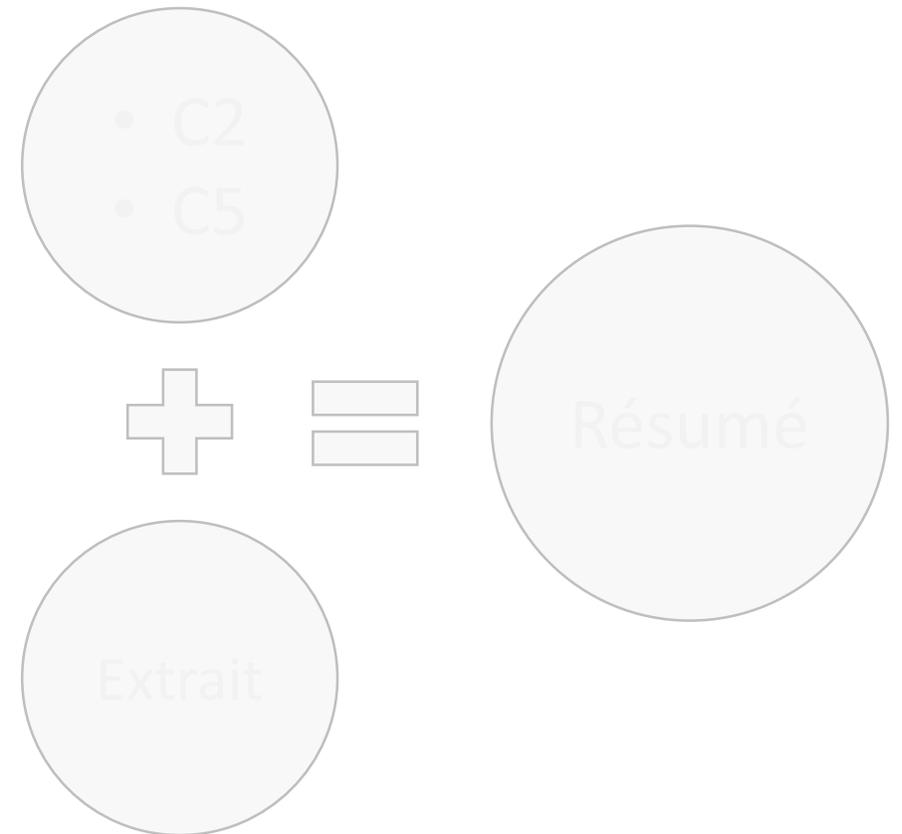
---

- *Moggi and Sabry [\*] used operational techniques to prove the safety of the full ST monad with typed references.*
- *Haskell's ST monad and its STRef values guarantees such memory safety [\*] by using the same type variable  $s$  to tag both the type of memory references ( $STRef\ s\ a$ ) and the type of computations using these references ( $ST\ s\ b$ ). Not only are ST and STRef both abstract type constructors, but runST, the function for running ST computations, has a rank-2 type that universally quantifies over the type variable  $s$  and prevents it from 'leaking'.*
- *The linguistic framework of Haskell is Moggi's monadic metalanguage, ml, which has served as the de facto standard for monadic languages [\*].*

# Exemple : amélioration du résumé

---

*We formalize the intended implementations as big-step operational semantics (which are referred to as dynamic semantics), then we prove type safety for three systems.*



# Exemple : leçons apprises

---

- **Other** : nous donne ce qui a été retenu.
- **Basis** : nous donne des exemples d'utilisation.
- **Contrast** : nous donne les avancements sur le même sujet.
- Extrait : ajoute des détails spécifiques.

# Expérimentation

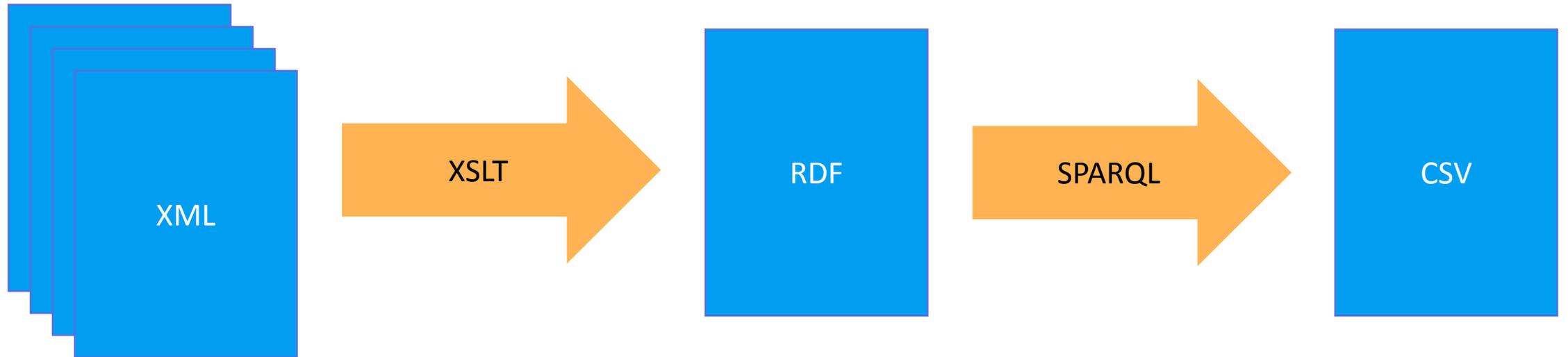
---

- Extended Semantic Web Conference 2014 (ESWC).
  - Semantic Publishing – Assessing the Quality of Scientific Output.
- Text Analysis Conference 2014 (TAC).
  - Biomedical Summarization.

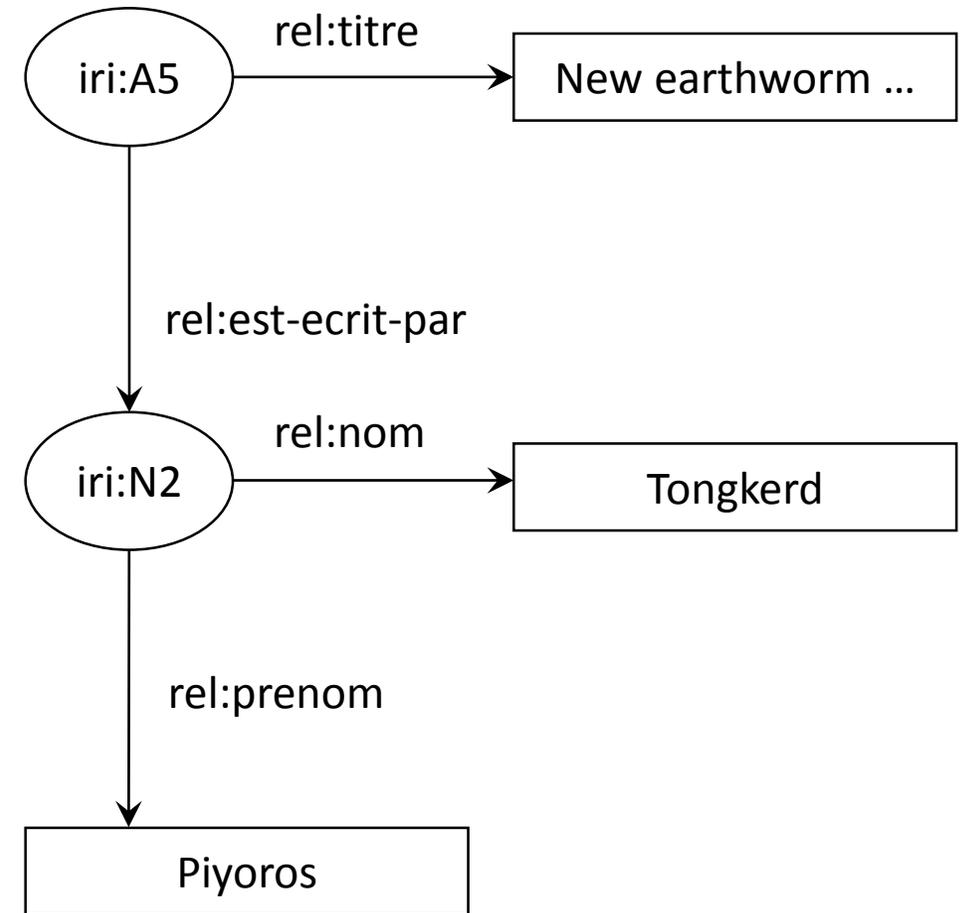
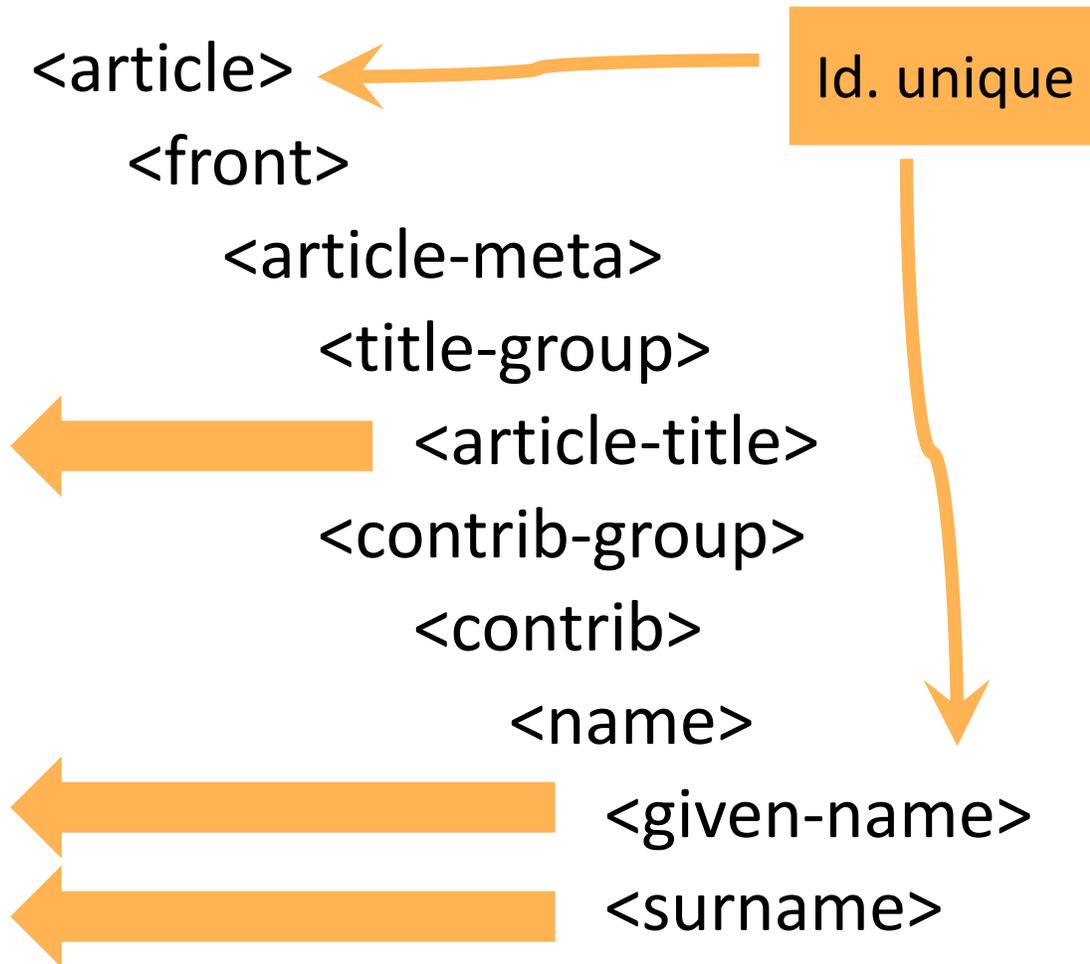
# Expérimentation : ESWC 2014

---

- Transformer des articles balisés (XML) en triplets RDF.
- Construire des requêtes SPARQL.



# Expérimentation : ESWC 2014



# Expérimentation : ESWC 2014

---

Identifier les articles cités par l'article X :

```
SELECT ?referenceiri ?doi ?pmid ?title
WHERE{
    FILTER ( ?X = "paper title here" ) .
    ?IRIciting    rel:article-titre ?X .
    ?IRIreference rel:citing      ?IRIciting .
    ?IRIreference rel:cited       ?referenceiri .
    ?referenceiri rel:article-titre ?title .
    OPTIONAL { ?referenceiri rel:doi   ?doi } .
    OPTIONAL { ?referenceiri rel:pmid ?pmid }
}
```

# Expérimentation : ESWC 2014

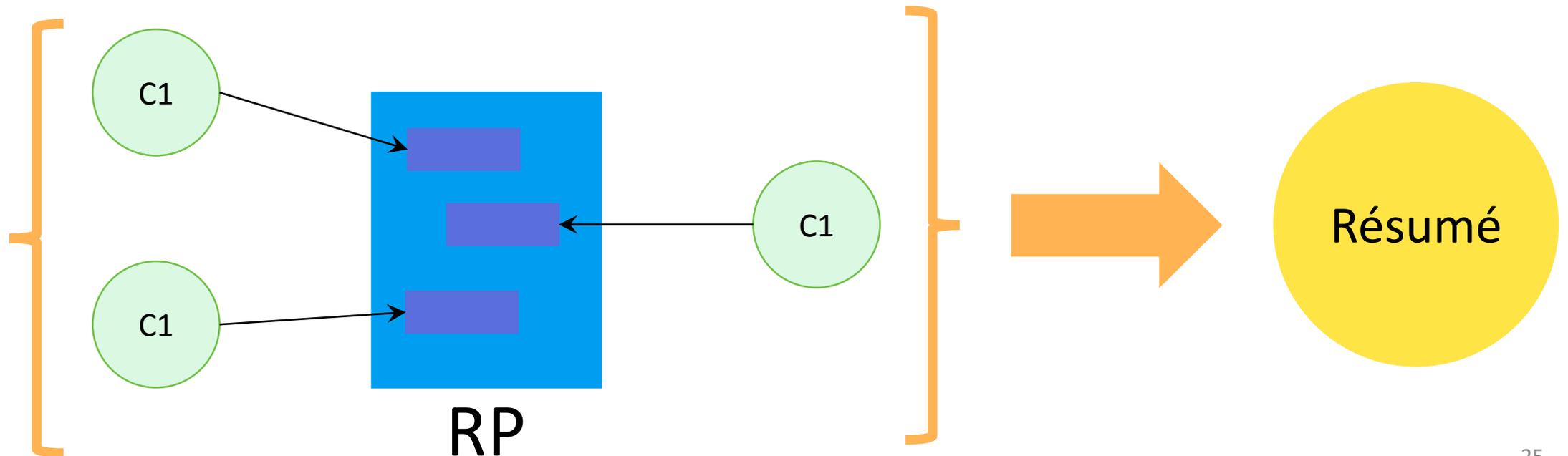
---

- Extraction de l'information facilitée par les balises.
- Requête SPARQL simple.

# Expérimentation : TAC 2014

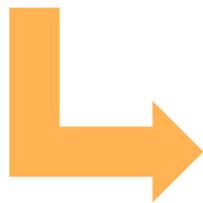
---

- Trouver le texte référé par une citance.
- Trouver la facette de ce texte.
- Construire un résumé à partir de cette information.



# Expérimentation : TAC 2014

*Moggi and Sabry [\*]*  
*prove*<sup>1</sup> *syntactic type*<sup>1</sup>  
*soundness for*  
*encapsulated lazy state*<sup>1</sup>



entrainement

## Methods

Prove	63
State	4
Type	35

## Results

Prove	74 + 1
State	43 + 1
Type	2 + 1

Methods :  $63 + 4 + 35 = 102$

Results :  $75 + 44 + 3 = 122$



classification

<sup>1</sup> Mots appartenant au Lexitrans (Patrick Drouin, 2010).

# Expérimentation : TAC 2014

---

- Baliser le texte.
- Entraîner pour les citances.
- Entraîner pour les phrases référées.
- Classer les citances.
- Classer les phrases de l'article de référence.
- Choisir la phrase référée qui a la même facette que la citance.
  - Mesure de similarité avec la citance et le résumé signalétique.
- Construction du résumé à l'aide des phrases de référence.

# Expérimentation : TAC 2014

---

- Difficile de baliser le texte.
- Le Lexitrans permet d'obtenir de bons résultats pour la classification.
- Malgré l'ensemble restreint de choix, il reste difficile de trouver la phrase de référence.

# Conclusion

---

- Les citances donnent de l'information nouvelle et utile pour mesurer l'impact d'un article.
- Il y a un intérêt pour les résumés d'articles scientifiques et leurs impacts: Comp Ling summarization.
- Nous avons proposé des techniques pour :
  - construire un résumé augmenté de son impact
  - détecter le rôle des articles citants.

# Conclusion

---

- Il existe déjà des classificateurs pour détecter les facettes des phrases d'un article.
- Nous avons des articles dont les citations ont été identifiées.
- Il reste à choisir les citations qui composeront le résumé :
  - Mesure de similarité.
- Trouver l'information complémentaire dans l'article de référence.