Extraction d'information

Philippe Langlais

felipe@iro.umontreal.ca

27 novembre 2018

École **IVADO/RALI**

https://ivado.ca/formations/ecoles/ecole-analyse-du-langage-naturel-2018/

Plan

Extraction d'information

Extraction d'information ouverte

Intermède

Structuration

Applications

Conclusion

Extraction d'information

What ?

https://en.wikipedia.org/wiki/Information_extraction

> *Information extraction. Information extraction (IE) is the task of automatically extracting structured information from unstructured and/or semi-structured machine-readable documents. In most of the cases this activity concerns processing human language texts by means of natural language processing (NLP).*

Extraire quoi ?

https://fr.wikipedia.org/wiki/Tigran_Hamasyan

[En 2009]$_{\text{TEMP}}$, il enregistre [Red Hail]$_{\text{ALBUM}}$, un album au carrefour du jazz, du metal et du folklore arménien, avec son nouveau quintet de jeunes musiciens [Aratta Rebirth]$_{\text{ORG}}$: [Areni Agbabian]$_{\text{PER}}$ (voc), [Ben Wendel]$_{\text{PER}}$ (ts), [Charles Altura]$_{\text{PER}}$ (g), [Sam Minaie]$_{\text{PER}}$ (b) et [Nate Wood]$_{\text{PER}}$ (d). Ils se produisent dans plusieurs grands festivals internationaux, de [Montréal]$_{\text{LOC}}$ à [Nice]$_{\text{LOC}}$ en passant par [Vienne]$_{\text{LOC}}$ ou [Rotterdam]$_{\text{LOC}}$ ([North Sea Jazz Festival]$_{\text{EVENT}}$).

- entités nommées, expressions temporelles, valeurs numériques
- relations entre entités

  ex : $\langle$Aratta Rebirth, se_produit, dans plusieurs grands festivals internationaux$\rangle$

Quelles relations ?

https://fr.wikipedia.org/wiki/Tigran_Hamasyan

[En 2009]$_{\text{TEMP}}$, il enregistre [Red Hail]$_{\text{ALBUM}}$, un album au carrefour du jazz, du metal et du folklore arménien, avec son nouveau quintet de jeunes musiciens [Aratta Rebirth]$_{\text{ORG}}$: [Areni Agbabian]$_{\text{PER}}$ (voc), [Ben Wendel]$_{\text{PER}}$ (ts), [Charles Altura]$_{\text{PER}}$ (g), [Sam Minaie]$_{\text{PER}}$ (b) et [Nate Wood]$_{\text{PER}}$ (d). Ils se produisent dans plusieurs grands festivals internationaux, de [Montréal]$_{\text{LOC}}$ à [Nice]$_{\text{LOC}}$ en passant par [Vienne]$_{\text{LOC}}$ ou [Rotterdam]$_{\text{LOC}}$ ([North Sea Jazz Festival]$_{\text{EVENT}}$).

- relations particulières : $\langle artiste \rangle$, *a_produit*, $\langle album \rangle$
  `(Tigran_Hamasyan,Red_Hail)`
- ou toutes les relations exprimées . . . ou le plus possible !

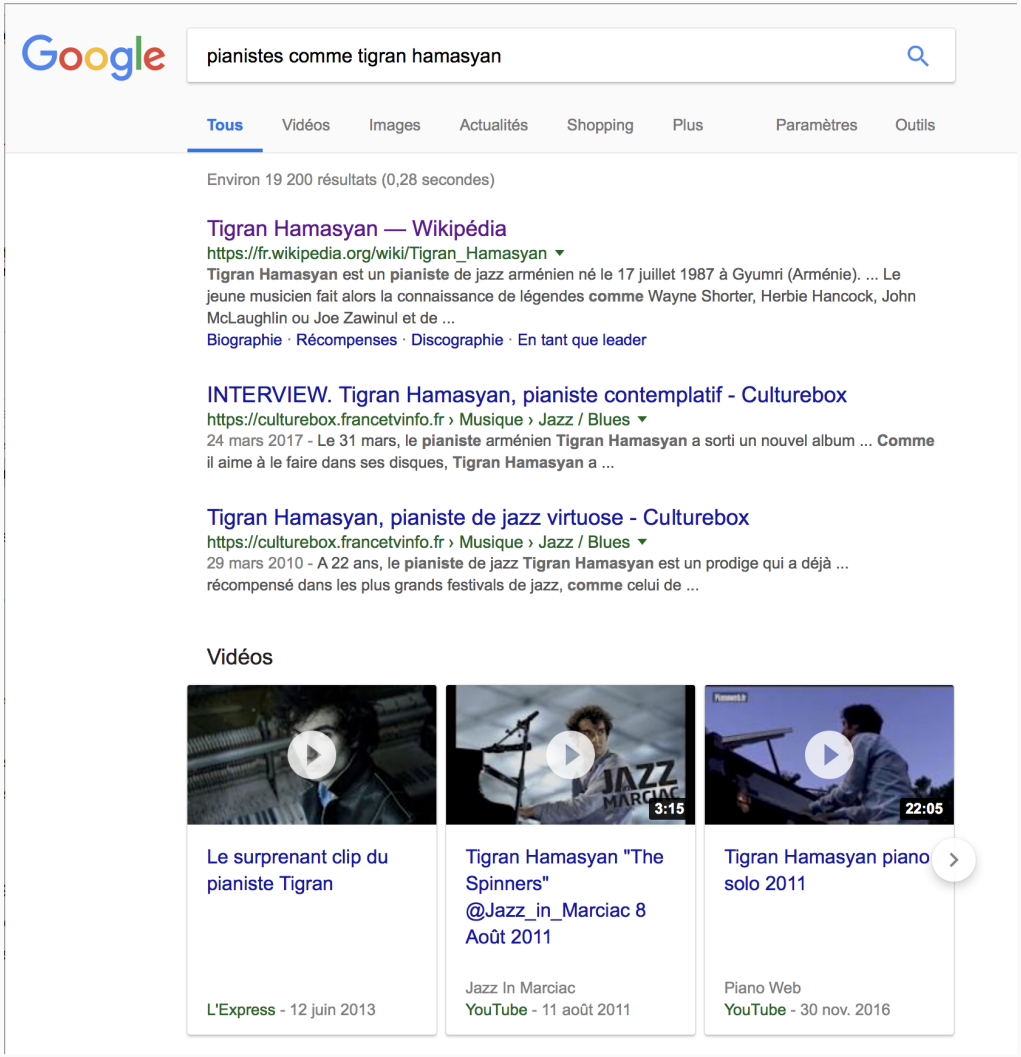
Dans quel but ?

Peupler une base de connaissances pour :

- une meilleure recherche d'information

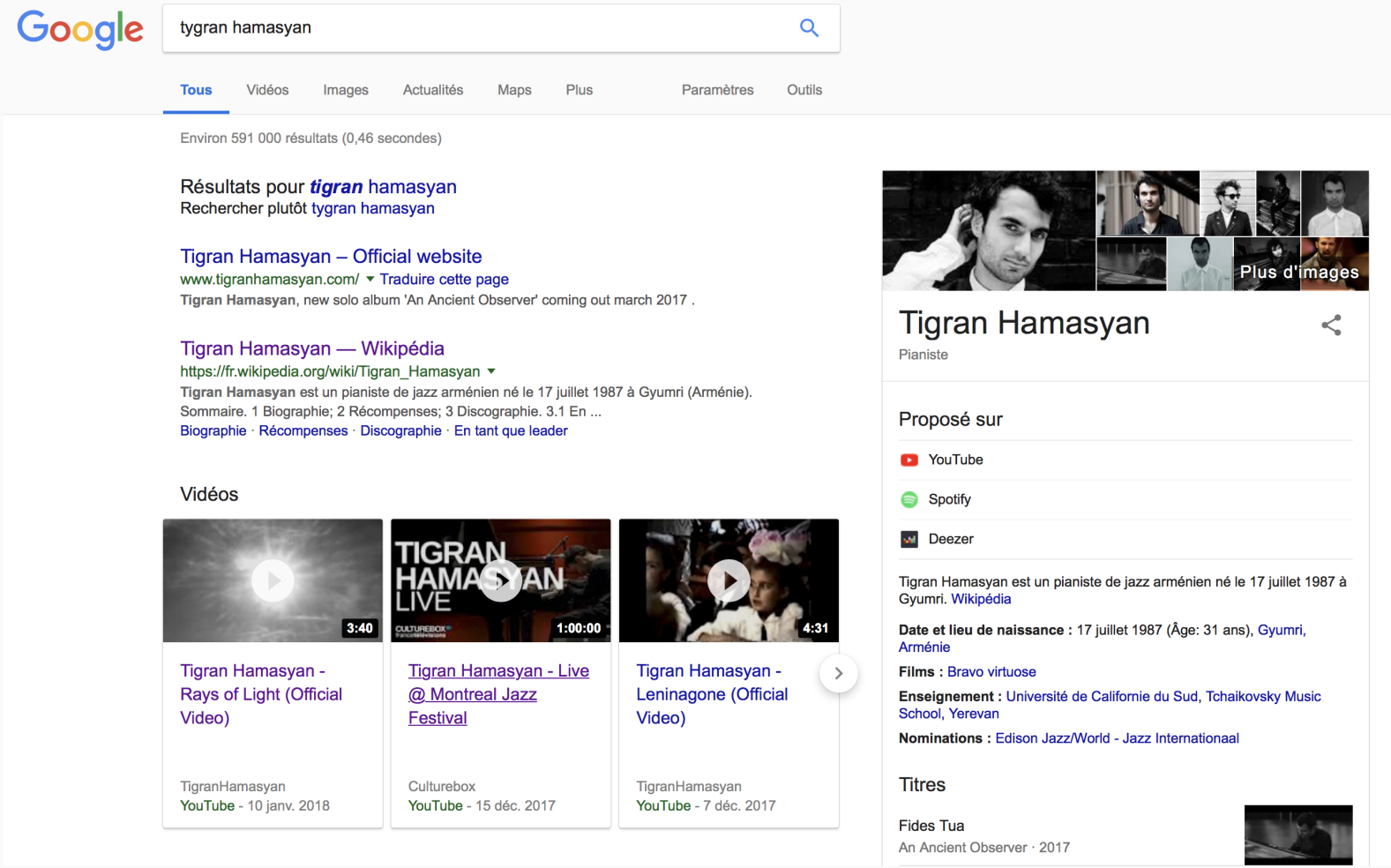
Meilleure RI

Pianistes comme Tigran Hamasyan



Meilleure RI

Tigran Hamasyan



Meilleure RI

Tigran Hamasyan

Pianiste

Proposé sur

- YouTube
- Spotify
- Deezer

Tigran Hamasyan est un pianiste de jazz arménien né le 17 juillet 1987 à Gyumri. Wikipédia

Date et lieu de naissance : 17 juillet 1987 (Âge: 31 ans), Gyumri, Arménie

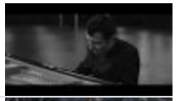
Films : Bravo virtuose

Enseignement : Université de Californie du Sud, Tchaikovsky Music School, Yerevan

Nominations : Edison Jazz/World - Jazz Internationaal

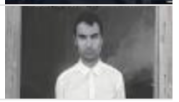
Titres

Fides Tua
An Ancient Observer · 2017



The Cave of Rebirth
An Ancient Observer · 2017

Ancient Observer
An Ancient Observer · 2017



Afficher 25 autres éléments

Albums

Afficher 3 autres éléments


Mockroot
2015


An Ancient Observer
2017

Shadow Theater
2013


Luys I Luso
2015


For Gyumri
2018

Recherches associées

Afficher 10 autres éléments


Lars Danielsson


Eivind Aarset


Arve Henriksen


Jan Bang

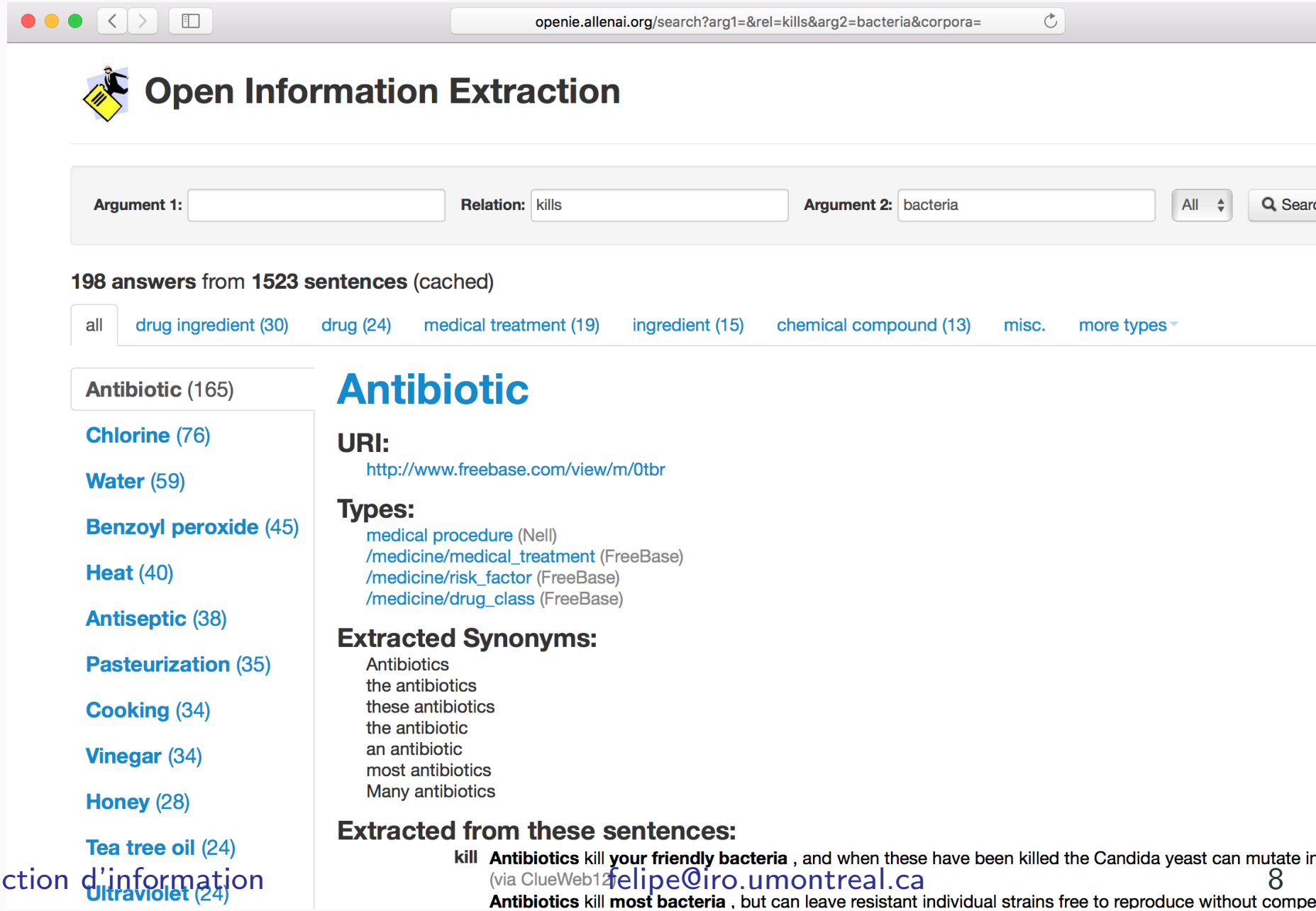

Avishai Cohen

Dans quel but ?

Peupler une base de connaissances pour :

- une meilleure recherche d'information
- outils de fouille

Outils de Fouille



Meilleure compréhension

Un sous-ensemble de *benchmarks* de question-réponse :

- SQUAD
- MS Marco
- NewsQA
- MC Test
- Cloze-style
- **RACE**
- CNN/Daily
- WDW (What did what)
- **AI2 Elementary school Science questions**
- triviaQA
- WikiHop

ARC (AI2 Reasoning Challenge) [Clark et al., 2018]

- 7787 questions à choix multiples en sciences naturelles :

  Easy set : 5197 questions

  Challenge set : 2590 questions qui résistent à des baselines basés sur la RI ou la co-occurrence

- questions pour étudiants de 8 à 13 ans

Which property of a mineral can be determined just by looking at it?

(A) `luster` (B) `mass` (C) `weight` (D) `hardness`

ARC (AI2 Reasoning Challenge) [Clark et al., 2018]

- 7787 questions à choix multiples en sciences naturelles :

  Easy set : 5197 questions

  Challenge set : 2590 questions qui résistent à des baselines basés sur la RI ou la co-occurrence

- questions pour étudiants de 8 à 13 ans

Which property of a mineral can be determined just by looking at it?

(A) luster (B) mass (C) weight (D) hardness

ARC (AI2 Reasoning Challenge) [Clark et al., 2018]

- 7787 questions à choix multiples en sciences naturelles :

  Easy set : 5197 questions

  Challenge set : 2590 questions qui résistent à des baselines basés sur la RI ou la co-occurrence

- questions pour étudiants de 8 à 13 ans

A student riding a bicycle observes that it moves faster on a smooth road than on a rough road. This happens because the smooth road has

(A) `less gravity` (B) `more gravity` (C) `less friction` (D) `more friction`

ARC (AI2 Reasoning Challenge) [Clark et al., 2018]

- 7787 questions à choix multiples en sciences naturelles :

  Easy set : 5197 questions

  Challenge set : 2590 questions qui résistent à des baselines basés sur la RI ou la co-occurrence

- questions pour étudiants de 8 à 13 ans

A student riding a bicycle observes that it moves faster on a smooth road than on a rough road. This happens because the smooth road has

(A) less gravity (B) more gravity (C) less friction (D) more friction

RACE [Lai et al., 2017]

- 28k passages, 100k questions posées par des enseignants de l'anglais langue seconde à des étudiants Chinois âgés de 12 à 18 ans
- deux niveaux : RACE-M (12-15 ans), RACE-H (15-18)
- vocabulaire restreint (apprenants de l'anglais)

- dataset : http://www.cs.cmu.edu/~glai1/data/race/
- code : https://github.com/qizhex/RACE_AR_baselines

Passage:
In a small village in England about 150 years ago, a mail coach was standing on the street. It didn't come to that village often. People had to pay a lot to get a letter. The person who sent the letter didn't have to pay the postage, while the receiver had to.
"Here's a letter for Miss Alice Brown," said the mailman.
" I'm Alice Brown," a girl of about 18 said in a low voice.
Alice looked at the envelope for a minute, and then handed it back to the mailman.
"I'm sorry I can't take it, I don't have enough money to pay it", she said.
A gentleman standing around were very sorry for her. Then he came up and paid the postage for her.
When the gentleman gave the letter to her, she said with a smile, " Thank you very much, This letter is from Tom. I'm going to marry him. He went to London to look for work. I've waited a long time for this letter, but now I don't need it, there is nothing in it."
"Really? How do you know that?" the gentleman said in surprise.
"He told me that he would put some signs on the envelope. Look, sir, this cross in the corner means that he is well and this circle means he has found work. That's good news."
The gentleman was Sir Rowland Hill. He didn't forgot Alice and her letter.
"The postage to be paid by the receiver has to be changed," he said to himself and had a good plan.
"The postage has to be much lower, what about a penny? And the person who sends the letter pays the postage. He has to buy a stamp and put it on the envelope." he said . The government accepted his plan. Then the first stamp was put out in 1840. It was called the "Penny Black". It had a picture of the Queen on it.

Questions:

1): The first postage stamp was made _.
A. in England B. in America C. by Alice D. in 1910

2): The girl handed the letter back to the mailman because _ .
A. she didn't know whose letter it was
B. she had no money to pay the postage
C. she received the letter but she didn't want to open it
D. she had already known what was written in the letter

3): We can know from Alice's words that _ .
A. Tom had told her what the signs meant before leaving
B. Alice was clever and could guess the meaning of the signs
C. Alice had put the signs on the envelope herself
D. Tom had put the signs as Alice had told him to

4): The idea of using stamps was thought of by _ .
A. the government
B. Sir Rowland Hill
C. Alice Brown
D. Tom

5): From the passage we know the high postage made _ .
A. people never send each other letters
B. lovers almost lose every touch with each other
C. people try their best to avoid paying it
D. receivers refuse to pay the coming letters

Answer: ADABC

Table 1: Sample reading comprehension problems from our dataset.

Histoire

- JASPER (1992) système Reuters pour les nouvelles financières
- MUC (1987-1998) Message Understanding Conference
  - Naval operations messages
  - Terrorism in Latin American countries
  - Joint ventures and microelectronics domain.
  - News articles on management changes.
  - Satellite launch reports.
- ACE Automatic Content Extraction Conference
- GATE (1995) Université de Sheffield
- OpenCalais (2008) Reuters (service web)
- TAC KBP (2009-2017). Text Analysis Conference — Knowledge Base Population

Extraction d'information

Par règles

Approche par règles

- des règles faites à la main pour capturer une information bien précise
- utilisée par exemple pour élaborer des wrappers

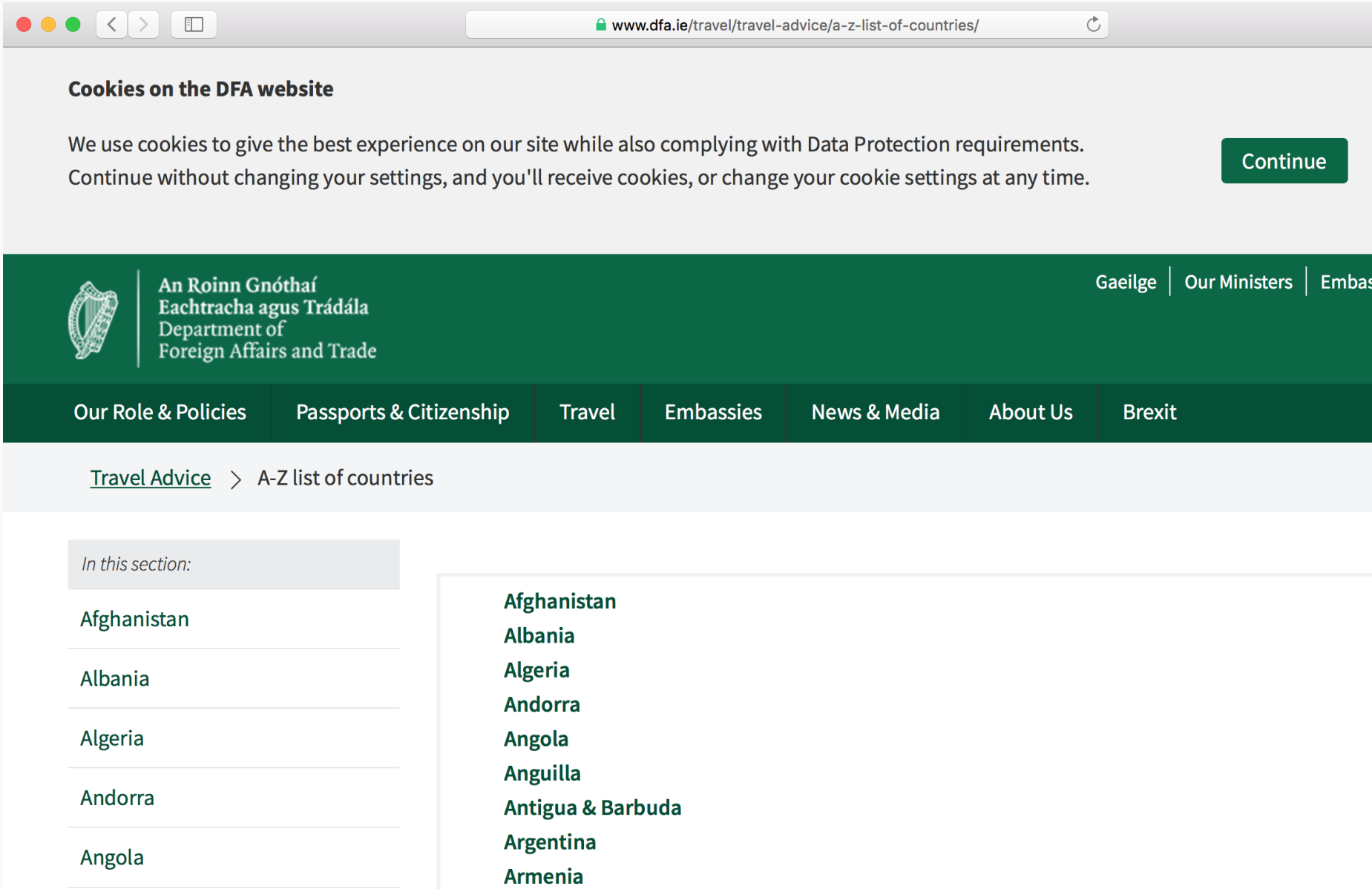
  in texte html

  out des entités (ou relations) d'intérêt
- possibilité d'apprendre les patrons (pour mieux généraliser)

Pro bonne précision

Cons très spécifique (à un domaine, à un type de marquage) $\Rightarrow$ faible rappel

Exemple (simpliste)

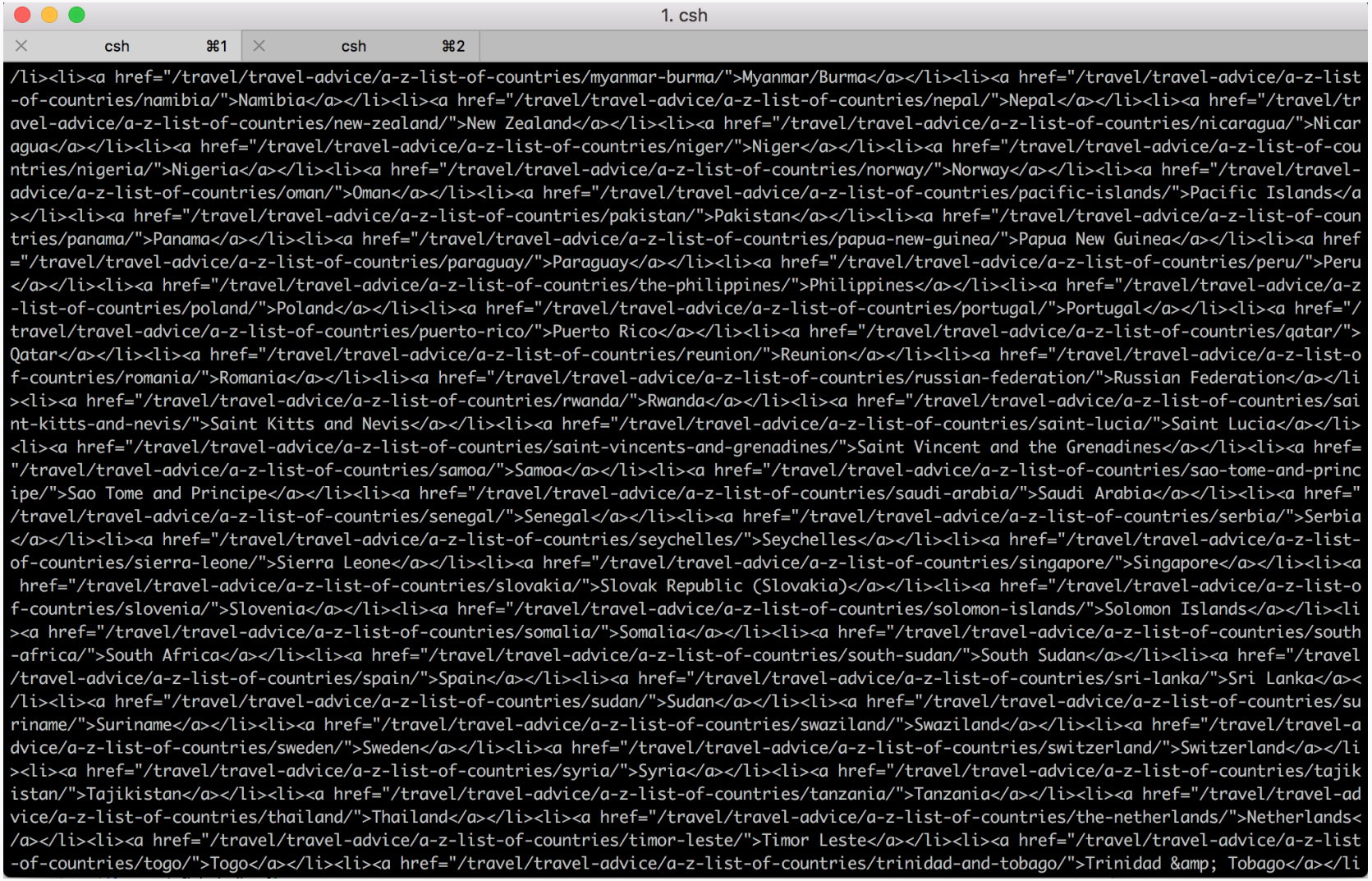


Exemple (simpliste)

```
654                         <!-- navigation object : 2017 Breadcrumbs --><li><a href="/travel/travel-advice/">Travel Advice</a></li><span class="breadcrumb_separat
655                     </ul>
656                 </nav>
657                 <!--** // breadcrumbs ** -->
658             </div>
659             <!--** // content-wrap **-->
660         </section>
661         <!--** // inner-page-breadcrumbs row -->
662         <!--** Main body row **-->
663         <section class="main-body main-body--general-content">
664             <div class="content-wrap">
665                 <aside class="main-body__side-bar">
666                     <!--^^ navigation object : 2017 Inner navigation ^^-->
667                     <h3 class="heading--side-bar">In this section:
668     <span class="inner-navigation__show-more__icon">
669         <svg class="dfa-svg-icon">
670             <use xmlns:xlink="http://www.w3.org/1999/xlink" xlink:href="#next_arrow"></use>
671         </svg>
672     </span>
673     <span class="inner-navigation__show-less__icon">
674         <svg class="dfa-svg-icon">
675             <use xmlns:xlink="http://www.w3.org/1999/xlink" xlink:href="#down_arrow"></use>
676         </svg>
677     </span>
678 </h3>
679 <ul class="inner-navigation"><li><a href="/travel/travel-advice/a-z-list-of-countries/afghanistan/">Afghanistan</a></li><li><a href="/travel/travel-advice/a-z-list-
680                     <!-- Script adds show more/less links to inner navs with dropdowns -->
681                     <script>
682                         (function addShowMoreLessLinks() {
683                           if ($('[class*="multilevel-linkul"]').length > 0) {
684                             var $mlUl = $('[class*="multilevel-linkul"]');
685                             /* Add show more. show less icons if there are nested links */
```

Exemple (simpliste)

```
more www.dfa.ie.html | grep Argentina
```



Exemple (simpliste)

```
[felipe,/Users/felipe/home/Papers/Talks/Ecole-IVADO-2018]more countries.html | sed -e 's/<li>/@/g' | tr '@' '\n' | grep a-z-list | head -n
                            <li class="en lang-switcher secondary-nav-link"><a href="/travel/travel-advice/a-z-list-of-countries/">Englis
English -->
                                    <li class="en lang-switcher secondary-nav-link"><a href="/travel/travel-advice/a-z-list-of-countries/
i> <!-- English -->
<a href="/travel/travel-advice/a-z-list-of-countries/afghanistan/">Afghanistan</a></li>
<a href="/travel/travel-advice/a-z-list-of-countries/albania/">Albania</a></li>
<a href="/travel/travel-advice/a-z-list-of-countries/algeria/">Algeria</a></li>
<a href="/travel/travel-advice/a-z-list-of-countries/andorra/">Andorra</a></li>
<a href="/travel/travel-advice/a-z-list-of-countries/angola/">Angola</a></li>
<a href="/travel/travel-advice/a-z-list-of-countries/anguilla/">Anguilla</a></li>
<a href="/travel/travel-advice/a-z-list-of-countries/antigua-and-barbuda/">Antigua &amp;amp; Barbuda</a></li>
<a href="/travel/travel-advice/a-z-list-of-countries/argentina/">Argentina</a></li>
[felipe,/Users/felipe/home/Papers/Talks/Ecole-IVADO-2018]
```

Exemple (simpliste)

```
[felipe,/Users/felipe/home/Papers/Talks/Ecole-IVADO-2018]more countries.html | sed -e 's/<li>/@/g' | tr '@' '\n' | grep a-z-list | grep '^<
<a href="/travel/travel-advice/a-z-list-of-countries/afghanistan/">Afghanistan</a></li>
<a href="/travel/travel-advice/a-z-list-of-countries/albania/">Albania</a></li>
<a href="/travel/travel-advice/a-z-list-of-countries/algeria/">Algeria</a></li>
<a href="/travel/travel-advice/a-z-list-of-countries/andorra/">Andorra</a></li>
<a href="/travel/travel-advice/a-z-list-of-countries/angola/">Angola</a></li>
<a href="/travel/travel-advice/a-z-list-of-countries/anguilla/">Anguilla</a></li>
<a href="/travel/travel-advice/a-z-list-of-countries/antigua-and-barbuda/">Antigua &amp;amp; Barbuda</a></li>
<a href="/travel/travel-advice/a-z-list-of-countries/argentina/">Argentina</a></li>
<a href="/travel/travel-advice/a-z-list-of-countries/armenia/">Armenia</a></li>
<a href="/travel/travel-advice/a-z-list-of-countries/australia/">Australia</a></li>
[felipe,/Users/felipe/home/Papers/Talks/Ecole-IVADO-2018]
```

Exemple (simpliste)

```
[felipe,/Users/felipe/home/Papers/Talks/Ecole-IVADO-2018]more countries.html | sed -e 's/<li>/@/g' | tr '@' '\n' | grep a-z-list | grep '^<
^>]*>\([^>]*\)<.*/\1/1' | head -n 20
Afghanistan
Albania
Algeria
Andorra
Angola
Anguilla
Antigua &amp;amp; Barbuda
Argentina
Armenia
Australia
Austria
Azerbaijan
Bahamas
Bahrain
Bangladesh
Barbados
Belarus
Belgium
Belize
Benin
[felipe,/Users/felipe/home/Papers/Talks/Ecole-IVADO-2018]
```

- on peut espérer apprendre une règle du genre :
  $\langle$ul$\rangle \ldots \langle$li$\rangle$ $\langle$ a href=”*countries*” $\rangle$ X $\langle$/a$\rangle$ $\langle$/li$\rangle$ $\ldots \langle$/ul$\rangle \Rightarrow$ X

Patrons lexicaux

- acronymes :

  `NP ([A-Z]+)`

⇒ `This is called Open Information retrieval (OIE), a task ...`

- hyperonymes **[Hearst, 1992]** :

  `NP such as NP`

  `NP is a NP`

  `NP, including {NP, }* {or | and} NP`

  `NP, especially NP`

  `...`

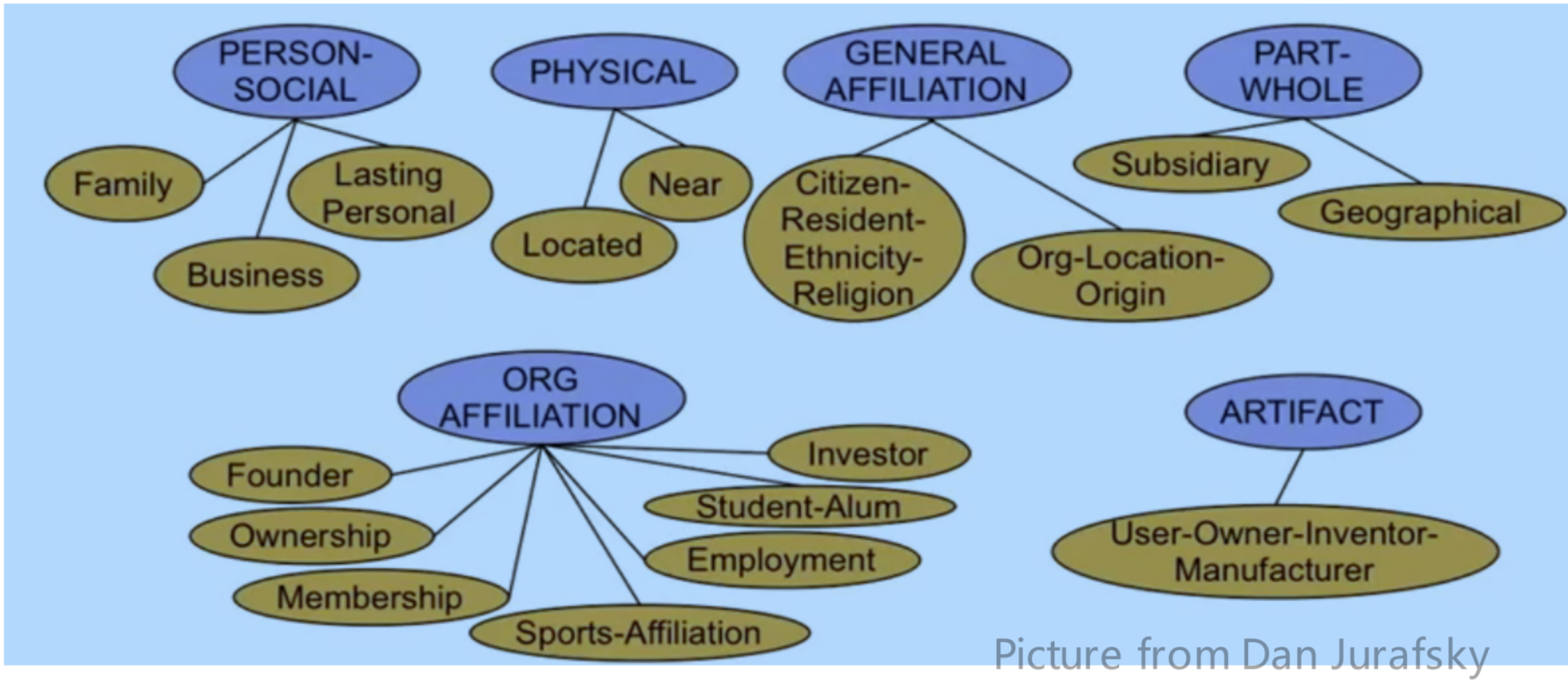
⇒
- riz, blé ∈ céréale
- nitrogène, phosphore ∈ substance

Extraction d'information

Supervisée

Automated Context Extraction (ACE)

- beaucoup de travaux dédiés à reconnaitre quelques dizaines de relations :



Picture from Dan Jurafsky

Automated Context Extraction (ACE)

[GuoDong et al., 2005]

Features	P	R	F
Words	69.2	23.7	35.3
+Entity Type	67.1	32.1	43.4
+Mention Level	67.1	33.0	44.2
+Overlap	57.4	40.9	47.8
+Chunking	61.5	46.5	53.0
+Dependency Tree	62.1	47.2	53.6
+Parse Tree	62.3	47.6	54.0
+Semantic Resources	63.1	49.5	55.5

Table 2: Contribution of different features over 43 relation subtypes in the test data

- requiert des centaines d'exemples d'entraînement pour chaque relation
- pour des performances finalement modestes

Extraction d'information

Semi-supervisée

DIPRE (Brin, 1998)

in : $\mathcal{D}$ une collection web de textes et S quelques exemples de paires d'entités validant la relation d'intérêt

Expérience

- recherche d'auteurs et leurs livres, à l'aide du *seed* S :

Isaac Asimov	The Robots of Dawn
David Brin	Startide Rising
James Gleick	Chaos : Making a New Science
William Shakespeare	The Comedy of Errors

- depuis une collection $\mathcal{D}$ de 24 millions de page Web (échantillonnées à chaque itération)

DIPRE (Brin, 1998)

in : $\mathcal{D}$ une collection web de textes et S quelques exemples de paires d'entités validant la relation d'intérêt

1. **Chercher les occurrences** des paires de S dans $\mathcal{D}$;
   pour chacune d'elle retenir un 7-tuple
   $\langle \text{arg}_1, \text{arg}_2, \text{order}, \text{url}, \text{prefix}, \text{middle}, \text{suffix} \rangle$
   - $\surd$ read The Adventures of Sherlock Holmes by Arthur Conan Doyle online or in your mail ...
   - $\surd$ I recommend you to read Startide by David Brin online at ...

DIPRE (Brin, 1998)

in : $\mathcal{D}$ une collection web de textes et S quelques exemples de paires d'entités validant la relation d'intérêt

1. **Chercher les occurrences** des paires de S dans $\mathcal{D}$;
   pour chacune d'elle retenir un 7-tuple
   $\langle \text{arg}_1, \text{arg}_2, \text{order}, \text{url}, \text{prefix}, \text{middle}, \text{suffix} \rangle$
   - $\langle$ACD, TASH, `false`, url_1, read, by, online or$\rangle$
   - $\langle$DB, Startide, `false`, url_2, you to read, by, online at$\rangle$

DIPRE (Brin, 1998)

in : $\mathcal{D}$ une collection web de textes et S quelques exemples de paires d'entités validant la relation d'intérêt

1. **Chercher les occurrences** des paires de S dans $\mathcal{D}$;
   pour chacune d'elle retenir un 7-tuple
   $\langle \text{arg}_1, \text{arg}_2, \text{order}, \text{url}, \text{prefix}, \text{middle}, \text{suffix} \rangle$
   - $\langle$ACD, TASH, `false`, url_1, read, by, online or$\rangle$
   - $\langle$DB, Startide, `false`, url_2, you to read, by, online at$\rangle$
2. **Induire des patrons** $\langle$order, urlprefix, prefix, middle, suffix$\rangle$
   - en groupant les 7-tuples qui ont le même `order` et `middle`
   - et en gardant les "parties correspondantes" (texte et url)
     - $\surd$ $\langle$`false`, url', read, by, online$\rangle$
       où url' est le préfixe commun à url_1 et url_2

Expérience

- les 3 patrons générés lors de la première itération :
  - ⟨ false, www.sff.net/locus/c., ⟨LI⟩ ⟨B⟩, ⟨/B⟩+ by, (⟩
    - ⟨LI⟩⟨B⟩title⟨/B⟩ by author (
  - ⟨ false, dns.city-net.com/~Imann/awards/hugos/1984.html ⟨i⟩, ⟨/i⟩ by, (⟩
    - ⟨i⟩ title ⟨/i⟩ by author (
  - ⟨ true, dolphin.upenn.edu/~dcummins/texts/sf-award.html, , ||, || (⟩
    - author||title|| (

DIPRE (Brin, 1998)

in : $\mathcal{D}$ une collection web de textes et S quelques exemples de paires d'entités validant la relation d'intérêt

1. **Chercher les occurrences** des paires de S dans $\mathcal{D}$;
   pour chacune d'elle retenir un 7-tuple
   $\langle \text{arg}_1, \text{arg}_2, \text{order}, \text{url}, \text{prefix}, \text{middle}, \text{suffix} \rangle$
   - $\langle$ACD, TASH, `false`, url_1, read, by, online or$\rangle$
   - $\langle$DB, Startide, `false`, url_2, you to read, by, online at$\rangle$
2. **Induire des patrons** $\langle \text{order}, \text{urlprefix}, \text{prefix}, \text{middle}, \text{suffix} \rangle$
   - en groupant les 7-tuples qui ont le même `order` et `middle`
   - et en gardant les "parties correspondantes" (texte et url)
     $\surd$ $\langle$`false`, url', read, by, online$\rangle$
     où url' est le préfixe commun à url_1 et url_2
3. **Appliquer les patrons** à la collection $\mathcal{D} \Rightarrow S$.
   Aller en 1 tant que nécessaire.

Expérience

- 3 itérations sur un échantillon de la collection :
  1. 199 occurrences, 3 patrons, 4047 paires
  2. sur 5M. de documents : 3972 occurrences, 105 patrons, 9369 paires
  3. sur un corpus dédié de 156k documents : 15257 paires

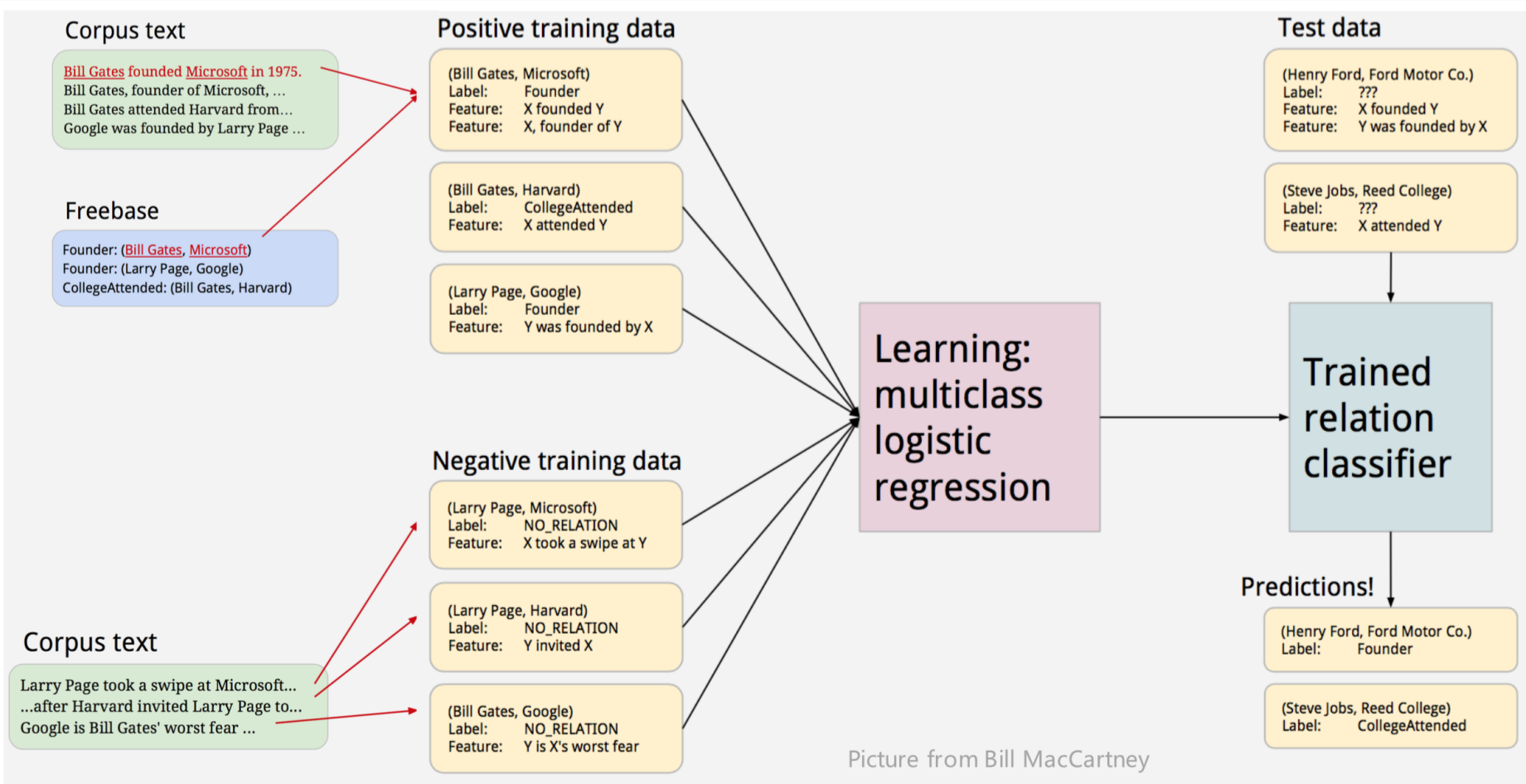
H. D. Everett	The Death-Mask and Other Ghosts
H. G. Wells	First Men in the Moon
H. G. Wells	Science Fiction: Volume 2
H. G. Wells	The First Men in the Moon
H. G. Wells	The Invisible Man
H. G. Wells	The Island of Dr. Moreau
H. G. Wells	The Science Fiction Volume 1
H. G. Wells	The Shape of Things to Come: The Ultimate Revolution
H. G. Wells	The Time Machine
H. G. Wells	The War of the Worlds
H. G. Wells	When the Sleeper Wakes
H. M. Hoover	Journey Through the Empty
H. P. Lovecraft & August Derleth	The Lurker at the Threshold
H. P. Lovecraft	At the Mountains of Madness and Other Tales of Terror
H. P. Lovecraft	The Case of Charles Dexter Ward
H. P. Lovecraft	The Doom That Came to Sarnath and Other Stories

Snowball [Agichtein and Gravano, 2000]

Très similaire à DIPRE, mais :

- patrons intégrant les types des entités en présence
  - $\surd$ ⟨org⟩'s headquarters in ⟨loc⟩
- mesure de confiance
  - des patrons détectés,
  - des paires initiales
  - **note :** requiert un corpus d'entraînement
- appliqué à l'extraction de paires (`organisation`, `siège social`)
  - $\surd$ (Microsoft, Redmond)
  - $\surd$ (Boeing, Seattle)

Supervision distante (Mintz et al. 2009)



Extraction d'information ouverte

Extraction d'information ouverte

Extracteurs

Extrateurs ouverts

- **Idée** : OIE = **pas** de liste de relations pré-établie
- Différents types d'extracteurs (disponibles souvent pour l'anglais)
  - POS-based : TextRunner, Woe^{pos}, ReVerb, Sonex
  - Dependency-based : Woe^{parse}, Ollie, TreeKernel, Patty, ClausIE, MinIE, Graphene, OpenIE
  - Semantic role labeling : Lund, SwiRL, Examplar
- Le premier : TextRunner **[Banko et al., 2007]**

Des sorties variées [Niklaus et al., 2018]

OLLIE:
(1) (Republican candidate Mitt Romney; will be elected President in; 2008)[enabler=If he w
(2) (Republican candidate Mitt Romney; will be elected; President)[enabler=If he wins five
(3) (Mitt Romney; be candidate of; Republican)
(4) (Mitt Romney; be candidate for; Republican)
(5) (he; wins; five key states)

ReVerb:
(6) (he; wins; five key states)
(7) (Republican candidate Mitt Romney; will be elected President in; 2008)

PredPatt:
(8) (he; wins; five key states)
(9) (Republican candidate Mitt Romney; will be elected President in; 2008)

ClausIE:
(10) (he; wins; five key states)
(11) (Republican candidate Mitt Romney; will be elected; President in 2008 If he wins five
(12) (Republican candidate Mitt Romney; will be elected; President in 2008)

OpenIE 5.0:
(13) (Republican candidate Mitt Romney; will be elected; President; T:in 2008)
(14) (he; wins; five key states)

Graphene:
(15) #1 CORE (Mitt Romney; will be elected; President)
("a) CONTEXT:NOUN_BASED Mitt Romney was a republican candidate .
("b) CONTEXT:TEMPORAL in 2008 .
("c) CONTEXT:CONDITION #3
("d) CONTEXT:NOUN_BASED #2
(16) #2 CORE (Mitt Romney; was; a republican candidate)
(17) #3 CONTEXT (he; wins; five key states)

Figure 1: Comparison of the output generated by different Open IE systems for th
wins five key states, Republican candidate Mitt Romney will be elected President

Woe [Wu and Weld, 2010]

- **idée :** utiliser les infobox de WIKIPEDIA comme supervision pour entraîner un extracteur

Pierre Lapointe

Pour les articles homonymes, voir Lapointe.

Pierre Lapointe (né le 23 mai 1981 à Alma, au Québec) est un auteur-compositeur-interprète canadien (québécois).

Son œuvre s'inscrit d'une part dans la tradition de la chanson francophone, qu'elle soit québécoise ou française, ce qui se vérifie dans des textes souvent très littéraires[non neutre]. Pierre Lapointe est également influencé par la musique pop qu'il entend utiliser pour renouveler la première citée[1]. Les arts graphiques contemporains, en particulier l'art numérique, sont très présents dans son univers via ses vidéoclips, et participent à la création d'un univers onirique et paradoxal, entre chansons mélancoliques et obscures, et scénographies colorées voire provocatrices.

Se définissant lui-même comme « chanteur populaire »[2,3], il s'est construit un personnage de dandy égocentrique de l'élite artistique montréalaise[4], lui permettant d'imprimer un décalage volontaire entre l'artiste sur scène et sa production largement biographique.

Ses disques enregistrent un succès critique et commercial au Canada[4].

Sommaire [masquer]
1 Biographie
1.1 Origines
1.2 Débuts
1.3 Les premiers succès
1.4 La confirmation
2 Analyse de l'œuvre
2.1 Thèmes
2.2 Influences
3 Discographie
3.1 Albums
3.2 Participations
4 Distinctions
5 Notes et références
5.1 Sources
6 Lien externe

Biographie [modifier | modifier le code]

Cette section **ne cite pas suffisamment ses sources**. Pour l'améliorer, ajouter en note des références vérifiables ou les modèles {{Référence nécessaire}} ou {{Référence souhaitée}} sur les passages nécessitant une source.

Pierre Lapointe

Pierre Lapointe lors d'un spectacle donné au Café théâtre le Côté-Cour de Jonquière (mars 2011).

Informations générales

Naissance	23 mai 1981 (33 ans) Alma, Québec, Canada
Activité principale	Chanteur, auteur-compositeur-interprète
Genre musical	Chanson française
Instruments	piano, voix
Années actives	Depuis 2000
Labels	Audiogram
Site officiel	www.pierrelapointe.com

- WOE est décrit en 2 saveurs :
  - WOE^{parse} qui utilise un analyseur en dépendance
  - WOE^{pos} qui utilise les POS (comme TEXTRUNNER)

Woe : 1 - distant supervision

Pour chaque article a de WIKIPEDIA, pour chaque paire attribut/valeur (t, v) dans l'infobox de a :

1. matcher dans les phrases de a le titre de a et v
   - présence de v :
     - correspondance exacte ou synonyme (via redirect, in-links) : *UK* vs *United Kingdom*
   - présence de *titre* :
     - correspondance exacte ou synonyme (redirects)
     - partial match : *Pierre* vs *Pierre Lapointe*
     - pronom le plus fréquent : *He* vs *Pierre Lapointe*
       sauf si le pronom le plus fréquent est it
2. filtrer les exemples positifs
   - une seule phrase dans a doit matcher v
   - titre et v dans la même clause, etc.

Woe : 1 - distant supervision

- pour 1M. d'articles WIKIPEDIA : 301 962 exemples positifs

Pierre Lapointe

Pierre Lapointe lors d'un spectacle donné au Café théâtre le Côté-Cour de Jonquière (mars 2011).

Informations générales

Naissance	23 mai 1981 (37 ans) Alma, Québec, Canada
Activité principale	Chanteur, auteur-compositeur-interprète
Genre musical	Chanson française
Instruments	piano, voix
Années actives	Depuis 2000
Labels	Audiogram
Site officiel	www.pierrelapointe.co m

- (activité principale, auteur-compositeur-interprète)
  ⚠ **Pierre Lapointe**, né le 23 mai 1981 à Alma, au Québec, est un **auteur-compositeur-interprète** canadien (québécois).
- (labels, audiogram)
  ⚠ En été 2003, **il** signe avec **Audiogram**.

Woe : 2 - extracteur (cas Woe^{parse})

1. analyser syntaxiquement (en dépendance) les phrases positives
   Pierre *was not born in* **Montreal**

   corePath *Pierre* $\xrightarrow{\text{nsubjpass}}$ *born* $\xleftarrow{\text{prep-in}}$ *Montréal*

   patrons *N* $\xrightarrow{\text{nsubjpass}}$ *V* $\xleftarrow{\text{prep}}$ *N*

2. une base de patrons avec leur fréquence
   - 15 333 patrons, 185 ayant une fréq. $\geq$100, 1929 $\geq$ 5
   - prob d'un patron est une fonction de la fréquence de ce patron
3. les dépendances non *core* (adjectivales, négations, etc.) sont ajoutées lors de l'extraction $\langle X, \textbf{was not } \textit{born} \textbf{ in}, Y \rangle$

Woe : Évaluation

- 900 phrases : 300 du WSJ, de Wikipedia et du Web
- **référence :** annotation manuelle de tous les triplets
  - $\text{Woe}^{parse} > \text{Woe}^{pos} > \text{TextRunner}$
  - nb moy. de tuples par phrase :
    - Woe^{parse} 1.42
    - Woe^{pos} 1.05
    - TextRunner 0.75
  - temps moyen pour traiter une phrase :
    - Woe^{parse} 0.679s
    - Woe^{pos} et TextRunner : 0.022s
      (Woe^{parse} est 30 fois plus lent)

ReVerb [Fader et al., 2011]

- 3 modules
  1. identification de relation verbale
  2. identification des arguments (NP) à gauche et à droite
  3. estimation de confiance (maxent)
- Évaluation
  - 500 pages Web retournées par `http://random.yahoo.com/bin/ryl` soumises à plusieurs extracteurs et dont les sorties sont évaluées manuellement

ReVerb : Extraction des relations

- contraintes **syntaxiques** :

  ```
  V | V P | V W* P
  V = verb particle? adv?
  W = (noun | adj | adv | pron | det)
  P = (prep | particle | inf. marker)
  ```

  invented (V), *located in* (VP), *has atomic weight of* (VW*P)

+ joindre deux relations adjacentes

ReVerb : Extraction des relations

- contraintes **syntaxiques** :

  `V | V P | V W* P`

  `V = verb particle? adv?`

  `W = (noun | adj | adv | pron | det)`

  `P = (prep | particle | inf. marker)`

  invented (V), *located in* (VP), *has atomic weight of* (VW*P)

+ joindre deux relations adjacentes

- contraintes **lexicales** :
  - éliminer les relations trop spécifiques

    is offering only modest greenhouse gas reduction targets at
  - une relation doit avoir plusieurs arguments dans une collection
    1. extraire toutes les relations et leurs arguments
    2. garder celles qui ont au moins 20 paires d'arguments différentes

ReVerb : Extraction des arguments

Pour chaque relation r :

- trouver x le NP le plus proche de r à gauche
  - qui n'est pas un pronom relatif, un adverbe WHO, ou `there`
- trouver y le NP le plus proche à droite de r

TOKS	Elle reprend ses études à l' Université McGill et obtient un doctorat en psychologie (1965) .
POS	CLS V DET NC P DET NC NPP CC V DET NC P NC PONCT NC PONCT PONCT
CHNK	B-NP B-VN B-NP I-NP B-PP B-NP I-NP I-NP B-COORD B-VN B-NP I-NP B-PP B-NP O B-NP O O

ReVerb : Extraction des arguments

Pour chaque relation *r* :

- trouver *x* le NP le plus proche de *r* à gauche
  - qui n'est pas un pronom relatif, un adverbe WHO, ou `there`
- trouver *y* le NP le plus proche à droite de *r*

TOKS	Elle reprend ses études à l' Université McGill et obtient un doctorat en psychologie (1965) .
POS	CLS V DET NC P DET NC NPP CC V DET NC P NC PONCT NC PONCT PONCT
CHNK	B-NP B-VN B-NP I-NP B-PP B-NP I-NP I-NP B-COORD B-VN B-NP I-NP B-PP B-NP O B-NP O O
EXTR	Elle == reprend == ses études

ReVerb : Extraction des arguments

Pour chaque relation *r* :

- trouver *x* le NP le plus proche de *r* à gauche
  - qui n'est pas un pronom relatif, un adverbe WHO, ou `there`
- trouver *y* le NP le plus proche à droite de *r*

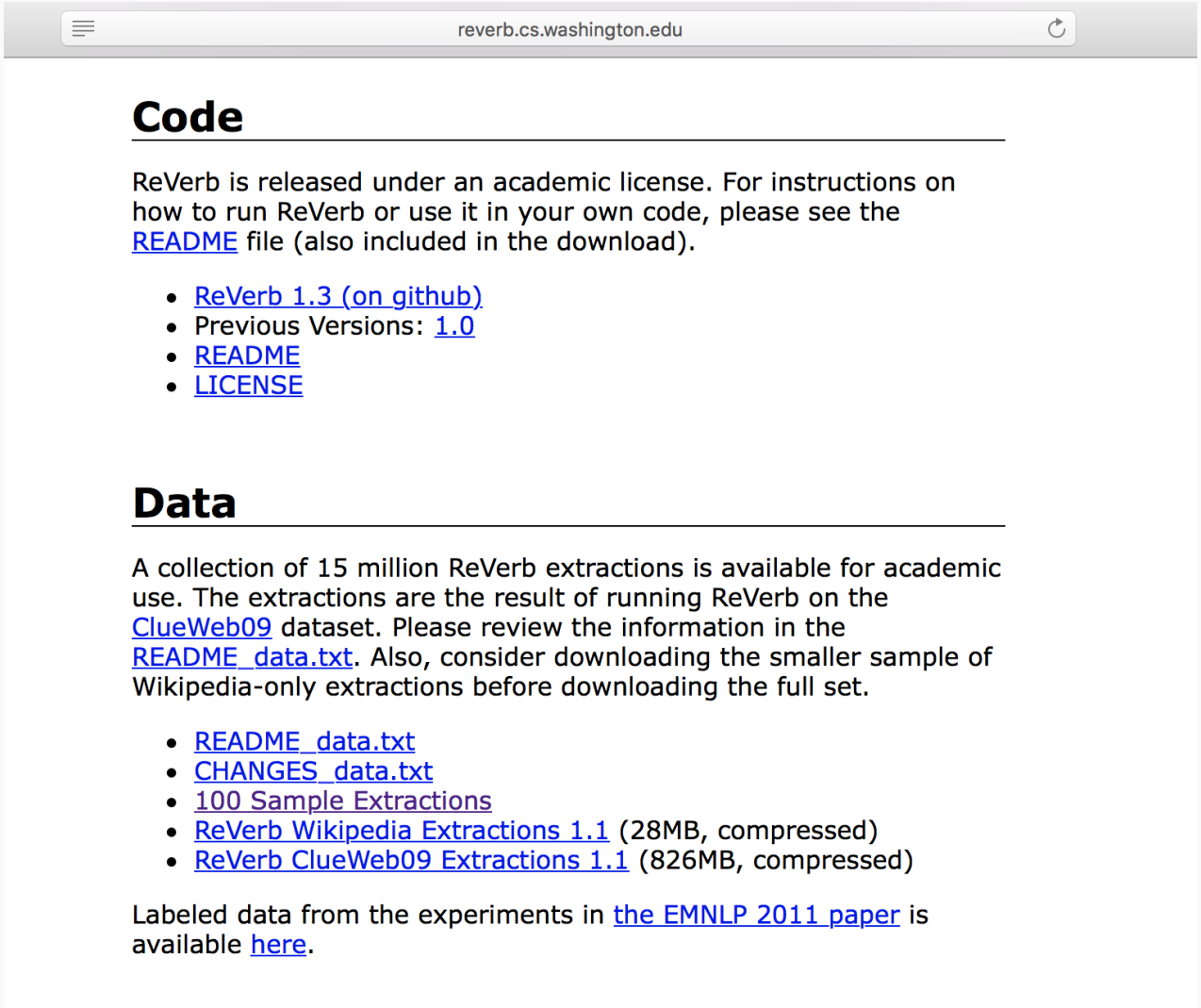
TOKS	Elle reprend ses études à l' Université McGill et obtient un doctorat en psychologie (1965) .
POS	CLS V DET NC P DET NC NPP CC V DET NC P NC PONCT NC PONCT PONCT
CHNK	B-NP B-VN B-NP I-NP B-PP B-NP I-NP I-NP B-COORD B-VN B-NP I-NP B-PP B-NP O B-NP O O
EXTR	Elle == reprend == ses études
EXTR	ses études == obtient == un doctorat

ReVerb : Estimateur de confiance

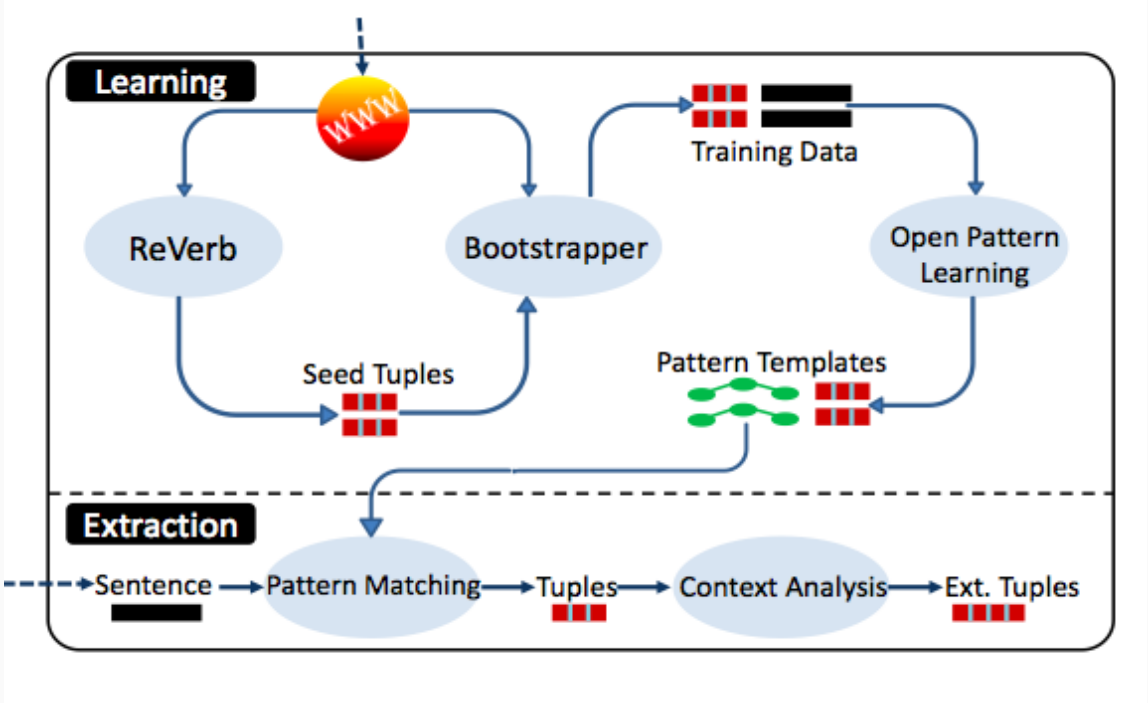
- **in :** (x, r, y) **out** : $p(\text{ok})$
- entraînement : référence manuelle de 1000 phrases
- 19 features qui ne dépendent pas d'une relation en particulier :

1.16	(x, r, y) couvre tous les mots de s
0.50	la dernière prép dans r est *for*
0.49	la dernière prép dans r est *on*
0.46	la dernière prép dans r est *of*
0.43	length(s) $\leq$ 10 mots
-0.93	conjonction de coord. à gauche de r dans s
	$\vdots$

ReVerb est disponible (pour l'anglais)



Ollie [Mausam et al., 2012]



- requiert des tuples extraits par ReVerb
- relations nominales : *Microsoft co-founder Bill Gates said...*
  - ⟨Bill Gates, is cofounder of, Microsoft⟩

Ollie : Apprentissage

- ∼110k tuples fiables extraits par ReVerb depuis ClueWeb :
  - freq≥2
  - args = noms propres
  - confiance élevée (selon ReVerb)
- 18M des phrases de ClueWeb contiennent ces tuples
- 4M une fois filtrées (mots de têtes des args reliés par un chemin de dépendances de long. au plus 4)
- **hyp** : ces phrases expriment des relations pertinentes
  vrai à 90% selon une éval. manuelle sur 100 phrases

Ollie : Apprentissage

Extraction Template	Open Pattern
1. (arg1; be {rel} {prep}; arg2)	{arg1} ↑nsubjpass↑ {rel:postag=VBN} ↓{prep_*}↓ {arg2}
2. (arg1; {rel}; arg2)	{arg1} ↑nsubj↑ {rel:postag=VBD} ↓dobj↓ {arg2}
3. (arg1; be {rel} by; arg2)	{arg1} ↑nsubjpass↑ {rel:postag=VBN} ↓agent↓ {arg2}
4. (arg1; be {rel} of; arg2)	{rel:postag=NN;**type=Person**} ↑nn↑ {arg1} ↓nn↓ {arg2}
5. (arg1; be {rel} {prep}; arg2)	{arg1} ↑nsubjpass↑ {slot:postag=VBN;**lex** ∈**announce**|**name**|**choose...**} ↓dobj↓ {rel:postag=NN} ↓{prep_*}↓ {arg2}

pris de **[Mausam et al., 2012]**

Analyse en dépendance des 4M de phrases retenues et extraction de patrons syntaxiques (droite) exprimant des relations (gauche) :

- des patrons purement syntaxiques (ex : 1 à 3)
  - généralisation aggressive (*on*, *on*, *for*, . . . = prep)

Ollie : Apprentissage

Extraction Template	Open Pattern
1. (arg1; be {rel} {prep}; arg2)	{arg1} ↑nsubjpass↑ {rel:postag=VBN} ↓{prep_*}↓ {arg2}
2. (arg1; {rel}; arg2)	{arg1} ↑nsubj↑ {rel:postag=VBD} ↓dobj↓ {arg2}
3. (arg1; be {rel} by; arg2)	{arg1} ↑nsubjpass↑ {rel:postag=VBN} ↓agent↓ {arg2}
4. (arg1; be {rel} of; arg2)	{rel:postag=NN;**type=Person**} ↑nn↑ {arg1} ↓nn↓ {arg2}
5. (arg1; be {rel} {prep}; arg2)	{arg1} ↑nsubjpass↑ {slot:postag=VBN;**lex** ∈**announce**|**name**|**choose...**} ↓dobj↓ {rel:postag=NN} ↓{prep_*}↓ {arg2}

pris de **[Mausam et al., 2012]**

Analyse en dépendance des 4M de phrases retenues et extraction de patrons syntaxiques (droite) exprimant des relations (gauche) :

- des patrons purement syntaxiques (ex : 1 à 3)
  - généralisation aggressive (*on*, *on*, *for*, . . . = prep)
- des patrons lexicalisés (ex : 5)

Ollie : Apprentissage

Extraction Template	Open Pattern
1. (arg1; be {rel} {prep}; arg2)	{arg1} ↑nsubjpass↑ {rel:postag=VBN} ↓{prep_*}↓ {arg2}
2. (arg1; {rel}; arg2)	{arg1} ↑nsubj↑ {rel:postag=VBD} ↓dobj↓ {arg2}
3. (arg1; be {rel} by; arg2)	{arg1} ↑nsubjpass↑ {rel:postag=VBN} ↓agent↓ {arg2}
4. (arg1; be {rel} of; arg2)	{rel:postag=NN;**type=Person**} ↑nn↑ {arg1} ↓nn↓ {arg2}
5. (arg1; be {rel} {prep}; arg2)	{arg1} ↑nsubjpass↑ {slot:postag=VBN;**lex** ∈**announce**|**name**|**choose...**} ↓dobj↓ {rel:postag=NN} ↓{prep_*}↓ {arg2}

pris de **[Mausam et al., 2012]**

Analyse en dépendance des 4M de phrases retenues et extraction de patrons syntaxiques (droite) exprimant des relations (gauche) :

- des patrons purement syntaxiques (ex : 1 à 3)
  - généralisation aggressive (*on*, *on*, *for*, . . . = prep)
- des patrons lexicalisés (ex : 5)
- des patrons typés (ex : 4)
  - généralisation des patrons lexicalisés grâce à WordNet

Ollie : Extraction

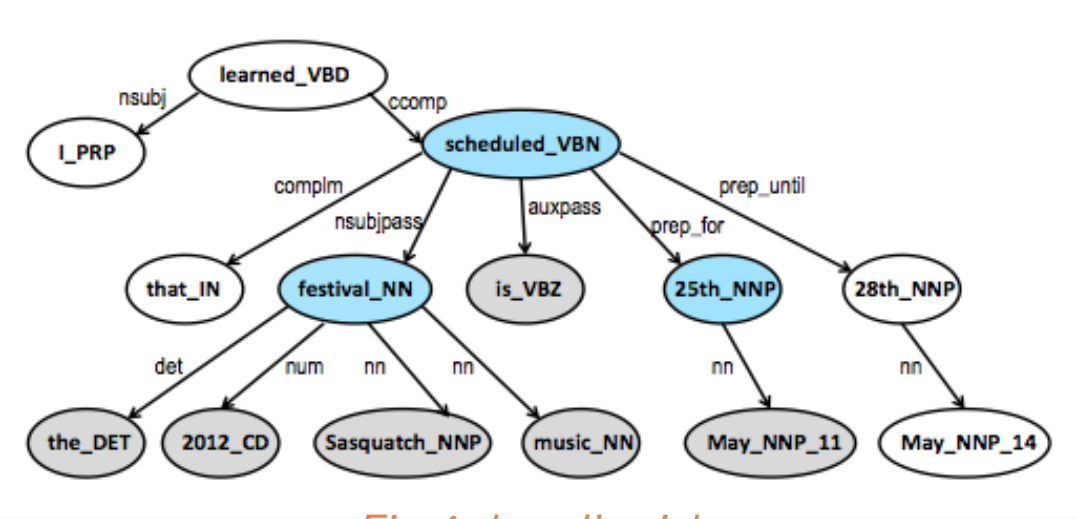


Fig 4 dans l'article

I learned that the 2012 Sasq. music festival is scheduled for May 25th until May 28th

- template : (arg ; be {rel} {prep} ; arg2)
- patron : {arg1} $\xrightarrow{nsubjpass}$ {rel :postag=VBN} $\xleftarrow{prep}$ {arg2}

1. match arg1=*festival* arg2=*25th* et rel=*scheduled*
   $\langle$*festival, be scheduled for, 25th*$\rangle$

Ollie : Extraction

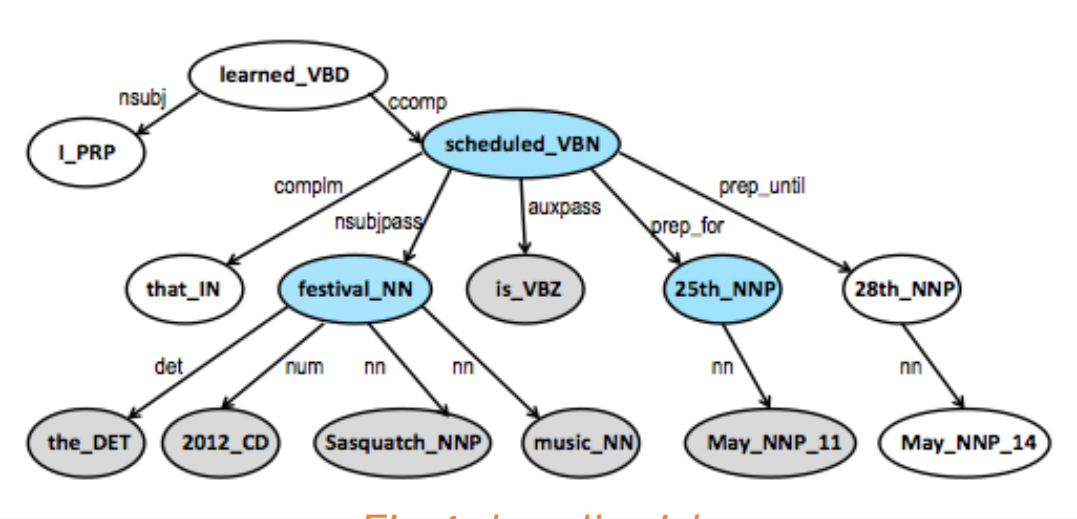


Fig 4 dans l'article

I learned that the 2012 Sasq. music festival is scheduled for May 25th until May 28th

- template : (arg ; be {rel} {prep} ; arg2)
- patron : {arg1} $\xrightarrow{nsubjpass}$ {rel :postag=VBN} $\xleftarrow{prep}$ {arg2}

1. match arg1=*festival* arg2=*25th* et rel=*scheduled*
   ⟨*festival*, *be scheduled for*, *25th*⟩
2. étendre les nœuds *args* et *rel* à certains mots gouvernés
   ⟨*the Sasquatch music festival*, *be scheduled for*, *May 25th*⟩

Ollie : Extraction

Ollie détecte 2 cas d'extractions non factuelles :

- *Early astronomers believed that the earth is the center of the universe*

R :	⟨the earth, be the center of, the universe⟩
O :	⟨the earth, be the center of, the universe⟩
	AttributedTo : believe ; Early astronomers

- *If he wins five states, Romney will be elected President*

R :	⟨Romney, will be elected, President⟩
O :	(⟨Romney, will be elected, President⟩
	ClausalModifier if ; he wins five key states)

OpenIE [Angeli et al., 2015]

- simplifier la structure syntaxique pour faciliter l'extraction

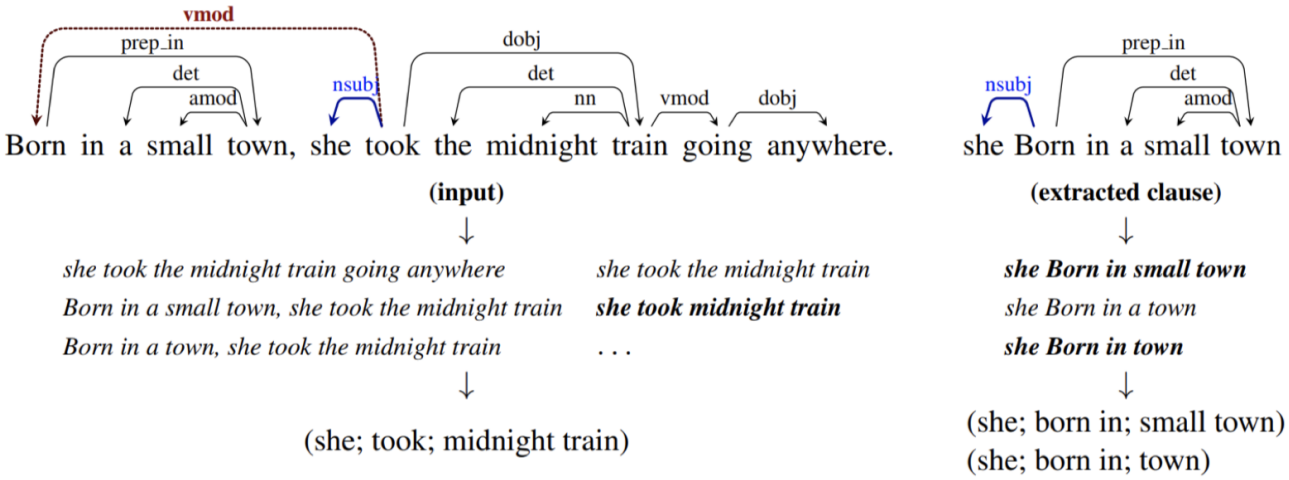


Figure 4: An illustration of Stanford Open IE's approach. From left to right, a sentence yields a number of independent clauses. From top to bottom, each clause produces a set of entailed shorter utterances, and segments the ones which match an atomic pattern into a relational triple (Angeli et al., 2015).

Pris de **[Niklaus et al., 2018]**

Buddleia

Triplet extractors

Chilly Gonzales (born Jason Charles Beck; 20 March 1972) is a Canadian musician who resided in Paris, France for several years, and now lives in Cologne, Germany.

Extract the triples | Extract with Ollie debug | Extract with Distylium debug

REVERB

20 March 1972)	is	a Canadian musician	0,56
a Canadian musician	resided in	Paris	0,79
France	now lives in	Cologne	0,87

OLLIE

20 March 1972	is a Canadian musician for	several years	0,86
20 March 1972	now lives in	Cologne	0,65
a Canadian musician	resided in	Paris	0,76
20 March 1972	is	a Canadian musician who resided in Paris	0,74

DISTYLIUM

a Canadian musician	resided in	Paris	56832,97
a Canadian musician	resided for	several years	56832,80
a Canadian musician	now lives in	Cologne	22315,53
Chilly Gonzales	is	a Canadian musician	10000,53
Paris	is	several years	38,63
a Canadian musician	resided	lives	24,87

Buddleia

Triplet extractors

Chilly Gonzales (born Jason Charles Beck; 20 March 1972) is a Canadian musician who resided in Paris, France for several years, and now lives in Cologne, Germany.

Extract the triples | Extract with Ollie debug | Extract with Distylium debug

CLAUSIE

Chilly Gonzales born Jason Charles Beck 20 March 1972	is	a Canadian musician	-254,85
a Canadian musician	resided	in Paris France for several years	-254,85
a Canadian musician	lives	in Cologne Germany	now -254,85

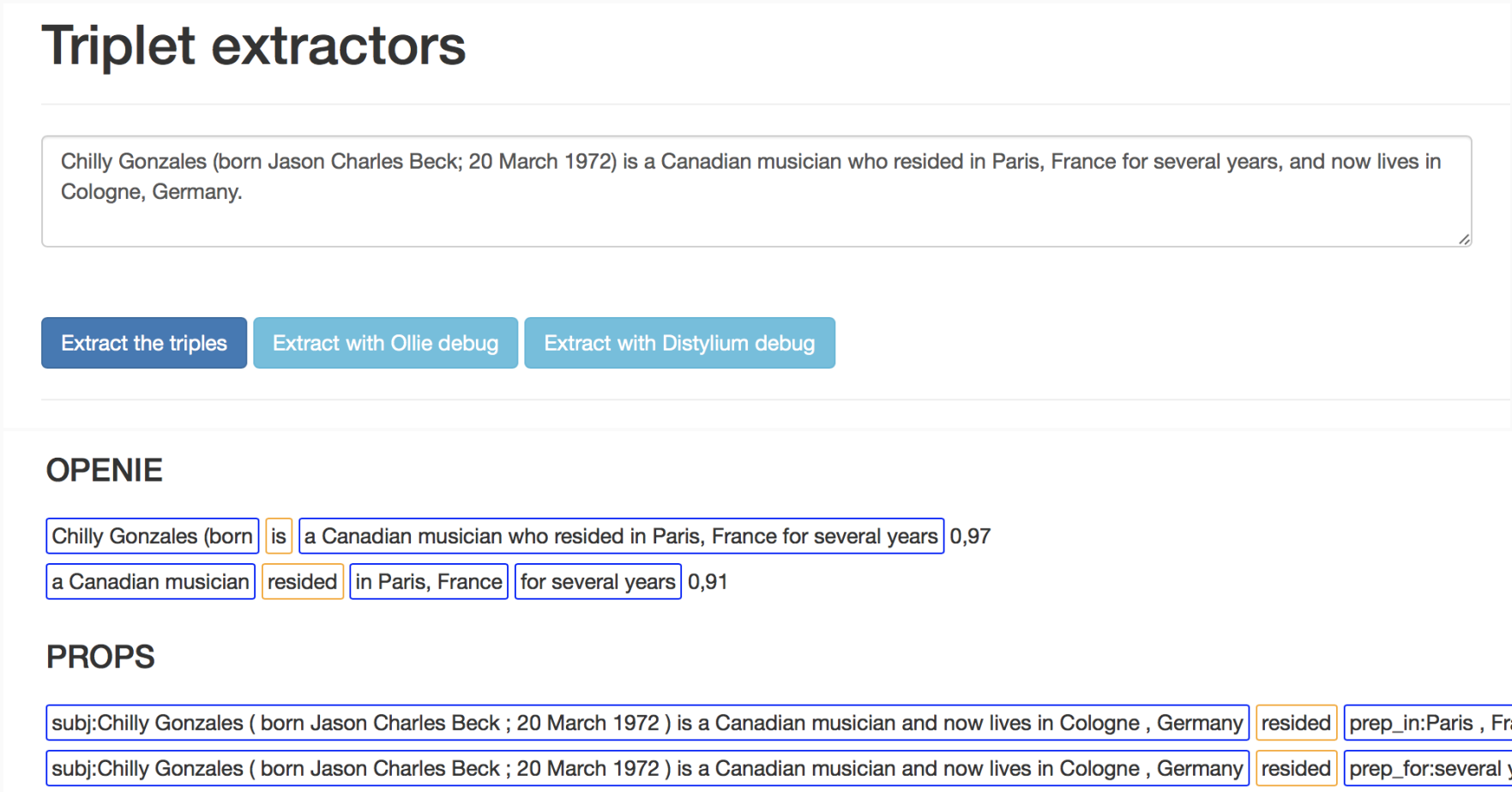
MINIE

Chilly Gonzales	is	Canadian musician	0,00
Canadian musician	resided in Paris for	QUANT_O_1 years | QUANT:[O_1=several]	0,00
Canadian musician	resided in	Paris	0,00
Canadian musician	lives in	Cologne	0,00
Paris	is in	France	0,00
Cologne	is in	Germany	0,00

STANFORD

Gonzales	is	Canadian	1,00

Buddleia


Triplet extractors
Chilly Gonzales (born Jason Charles Beck; 20 March 1972) is a Canadian musician who resided in Paris, France for several years, and now lives in Cologne, Germany.
Extract the triples
Extract with Ollie debug
Extract with Distylium debug
OPENIE
Chilly Gonzales (born
is
a Canadian musician who resided in Paris, France for several years
0,97
a Canadian musician
resided
in Paris, France
for several years
0,91
PROPS
subj:Chilly Gonzales (born Jason Charles Beck ; 20 March 1972) is a Canadian musician and now lives in Cologne , Germany
resided
prep_in:Paris , France
subj:Chilly Gonzales (born Jason Charles Beck ; 20 March 1972) is a Canadian musician and now lives in Cologne , Germany
resided
prep_for:several years

Extraction d'information ouverte

Limites

Analyse

- OLLIE et REVERB sur l'article WIKIPEDIA : Quebec

Article Talk Read View source View history Search

Quebec

From Wikipedia, the free encyclopedia

Coordinates: 53°N 07

This article is about the Canadian province. For the province's capital city, see Quebec City. For other uses, see Quebec (disambiguation).

Quebec (/kwɨˈbɛk/ or /keɪˈbɛk/; French: *Québec* [kebɛk] (listen))[7] is a province in east-central Canada.[8][9] It is the only Canadian province that has a predominantly French-speaking population, and the only one to have French as its sole provincial official language.

Quebec is Canada's largest province by area and its second-largest administrative division; only the territory of Nunavut is larger. It is bordered to the west by the province of Ontario, James Bay, and Hudson Bay; to the north by Hudson Strait and Ungava Bay; to the east by the Gulf of Saint Lawrence and the province of Newfoundland and Labrador; it is bordered on the south by the province of New Brunswick and the U.S. states of Maine, New Hampshire, Vermont, and New York. It also shares maritime borders with Nunavut, Prince Edward Island, and Nova Scotia.

Quebec
***Québec* (French)**

Flag Coat of arms

- 171 sentences : OLLIE a identifié 445 tuples, REVERB 217
- évaluation des 100 triplets les mieux notés

Coréférence : `the X`

- environ 50% des tuples

Jardin zoologique du Québec , reopened in 2002 after two years of restorations but closed in 2006 after a political decision . It featured 750 specimens of 300 different species of animals . **The zoo specialized in winged fauna and garden themes** , but also presented several species of mammals .

$$\langle \text{The zoo}, \text{specialized in}, \text{winged fauna and garden themes} \rangle$$

Coréférence : `the X`

- environ 50% des tuples

Jardin zoologique du Québec , reopened in 2002 after two years of restorations but closed in 2006 after a political decision . It featured 750 specimens of 300 different species of animals . **The zoo specialized in winged fauna and garden themes** , but also presented several species of mammals .

$$\langle \text{The zoo}, \text{specialized in}, \text{winged fauna and garden themes} \rangle$$

Coréférence : `the X`

- environ 50% des tuples

Porte St-Louis and Porte St-Jean are the main gates through the walls from the modern section of downtown ; the Kent Gate was a gift to the province from Queen Victoria and **the foundation stone was laid by the Queen 's daughter** , Princess Louise , Marchioness of Lorne , on June 11 , 1879 .

⟨the foundation stone, was laid by, the Queen 's daughter⟩

Coréférence : `the X`

- environ 50% des tuples

Porte St-Louis and Porte St-Jean are the main gates through the walls from the modern section of downtown ; the Kent Gate was a gift to the province from Queen Victoria and **the foundation stone was laid by the Queen 's daughter** , Princess Louise , Marchioness of Lorne , on June 11 , 1879 .

⟨the foundation stone, was laid by, the Queen 's daughter⟩

Coréférence : pronoms

Quebec City 's skyline is dominated by the massive Château Frontenac Hotel , perched on top of Cap-Diamant . **It was designed by architect Bruce Price** , as one of a series of " château " style hotels built for the Canadian Pacific Railway company .

$$\langle \text{It}, \text{was designed by}, \text{architect Bruce Price} \rangle$$

Coréférence : pronoms

Quebec City 's skyline is dominated by the massive Château Frontenac Hotel , perched on top of Cap-Diamant . **It was designed by architect Bruce Price** , as one of a series of " château " style hotels built for the Canadian Pacific Railway company .

$$\langle \text{It}, \text{was designed by}, \text{architect Bruce Price} \rangle$$

D'autres problèmes

- tuples non. informatifs :
  $\langle$The town, distinguished, itself$\rangle$
  $\langle$One-quarter of the people, were members of, religious$\rangle$
  $\langle$The team, has, size league titles$\rangle$

- syntaxe compliquée

The English-speaking community peaked in relative terms during the 1860s , when 40 % of Quebec City 's residents were Anglophone .

$\langle$The English-speaking community, peaked in, relative terms$\rangle$

Et pourtant ça marche !

- si appliqué à de larges collections (M. de phrases)
- en appliquant des filtres agressifs :
  tuples vu au moins n fois, score de confiance élevé, etc.

⟨Quebec, was founded by, Samuel de Champlain⟩
⟨Quebec City, is located on, the north bank of the Saint Lawrence River⟩
⟨Quebec City, is an important hub in, the province 's autoroute system⟩
⟨Quebec City, is located in, the Saint Lawrence River valley⟩
⟨Quebec City, was struck by, the 1925 Charlevoix-Kamouraska earthquake

Intermède

Intermède

deFacto

deFacto

1. extraction de tuples depuis 31M de phrases de Wikipedia-fr (2014) à l'aide d'une adaptation de ReVerb au français
   48h plus tard : 20M tuples une fois filtrés

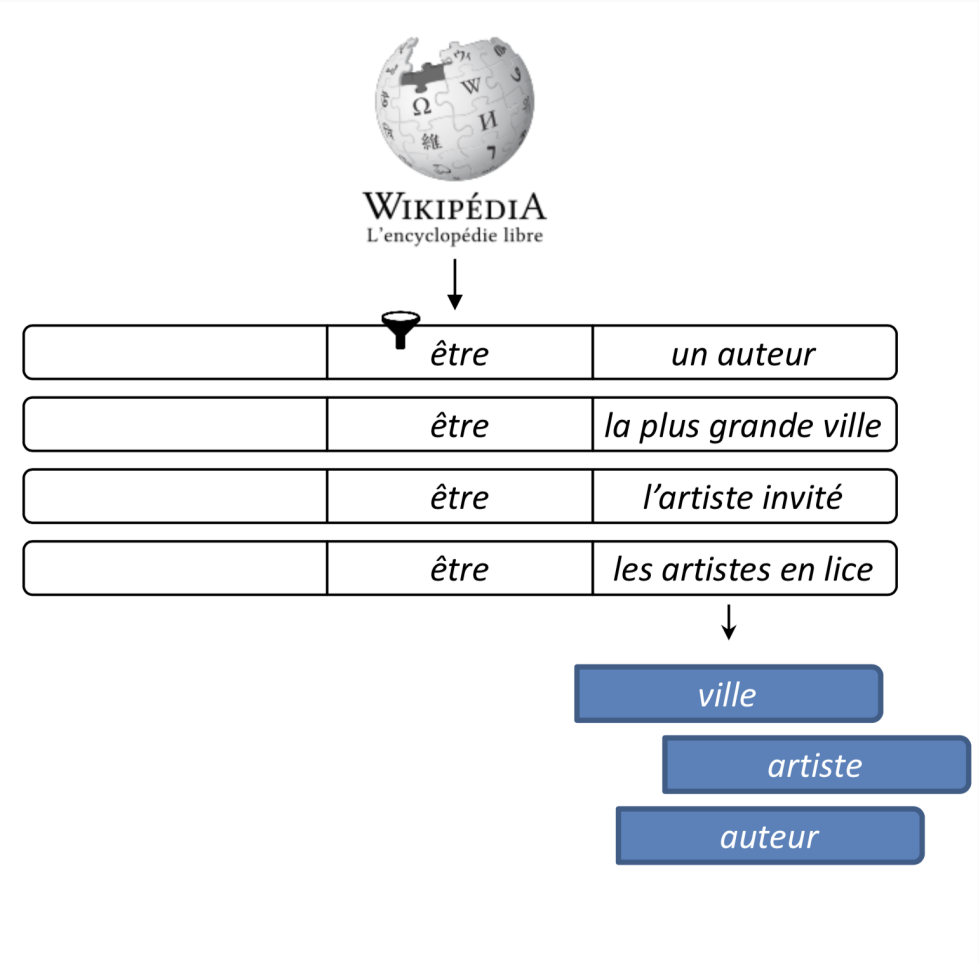
Extraction des tuples

...
CANO l' activité biologique == se dérouler dans == le lac
CANO les sédiments == être appeler == relargage
CANO sédiments == jouer == un rôle majeur
CANO l' eau == recouvrer == les sédiments
CANO les sédiments == contenir de == l' oxygène
CANO les échanges de P == être généralement == unidirectionnels de l' eau
CANO les conditions anoxiques == modifier == la nature des échanges chimiques
CANO la décomposition == anaérobie de == la matière organique
CANO La diminution du potentiel d' oxydoréduction == découler de == cette anoxie
CANO cette anoxie == favoriser == la remise
CANO biologiques == contribuer à == la libération du P
CANO Cette combinaison de facteurs == faire en == sorte
CANO les sédiments == ne pas être uniquement == un lieu d' accumulation du P
CANO le P == pouvoir être == maintes reprises
CANO Le lac Nairne == être affecter par == des épisodes de floraison de cyanobactéries annuellement
CANO des floraisons == être observer sur == toute la superficie du lac
CANO Certains épisodes == survenir en == été
CANO Ces floraisons massives == être == un symptôme de l' eutrophisation du lac
CANO Le lac Nairne == être à == coeur de la municipalité de Saint-Aimé-des-Lacs
CANO un plan d' action == être mettre sur == pied
CANO plusieurs actions == être entreprendre comme == la caractérisation

deFacto

1. extraction de tuples depuis 31M de phrases de Wikipedia-fr (2014) à l'aide d'une adaptation de ReVerb au français
2. arg_2s les plus fréquents de la relation *être* promus catégories

Sélection de catégories

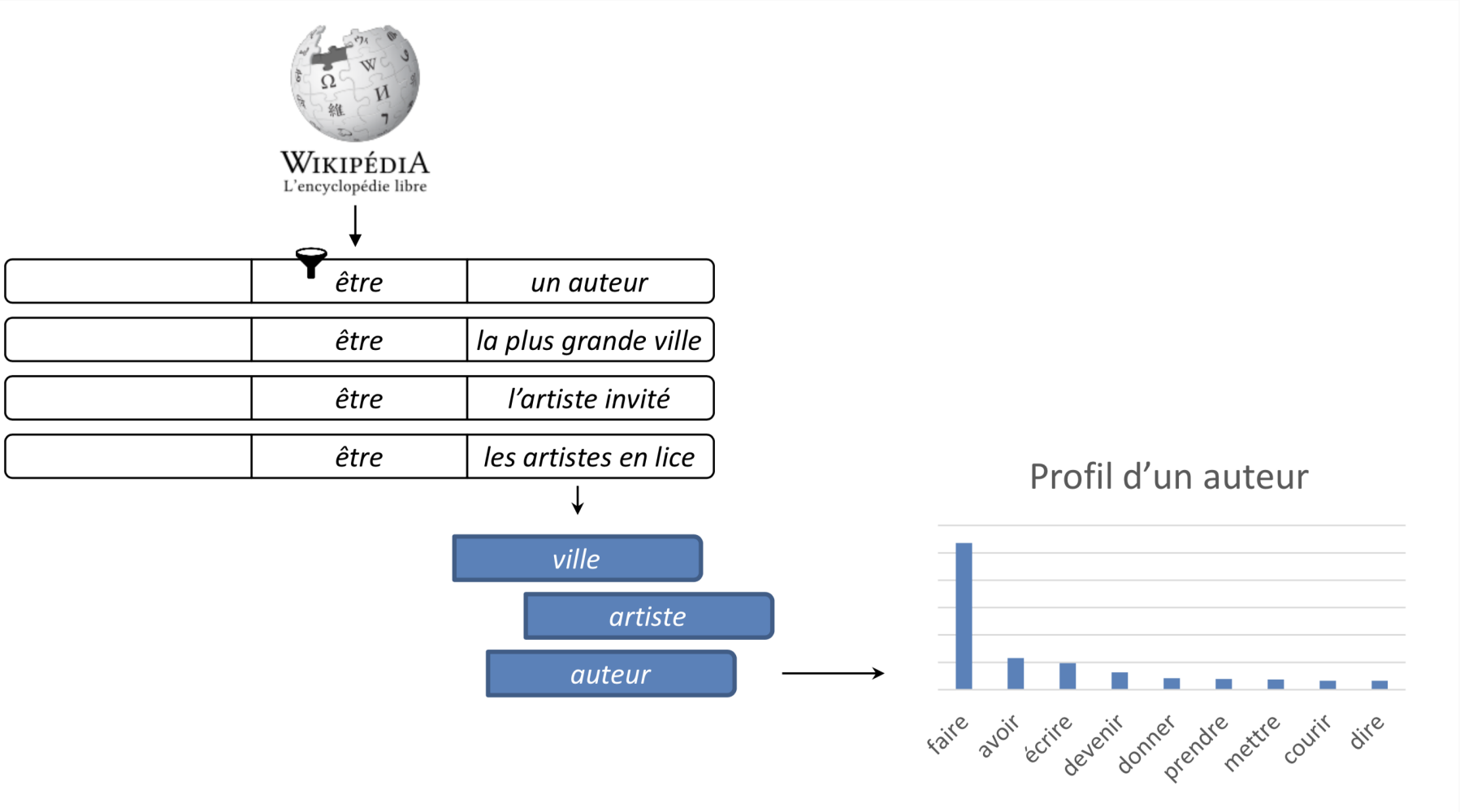


1. extraction de tuples depuis 31M de phrases de WIKIPEDIA-FR (2014) à l'aide d'une adaptation de REVERB au français
2. arg_2s les plus fréquents de la relation *être* promus catégories

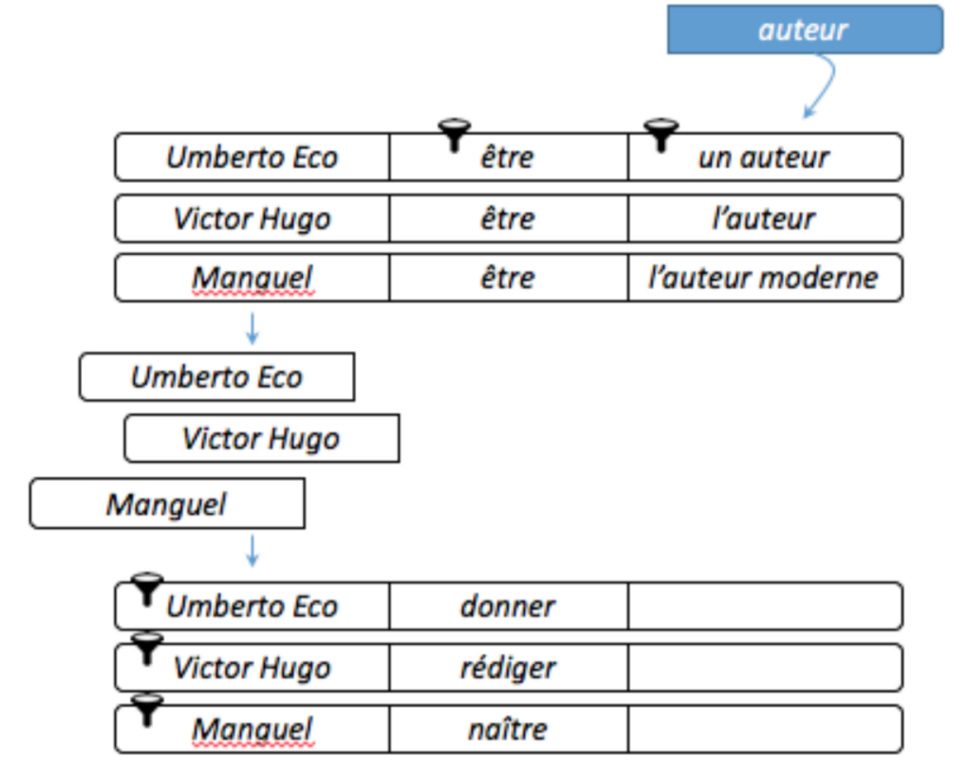
deFacto

1. extraction de tuples depuis 31M de phrases de Wikipedia-fr (2014) à l'aide d'une adaptation de ReVerb au français
2. arg_2s les plus fréquents de la relation *être* promus catégories
3. calcul du **profil relationnel** de chaque catégorie : les relations les plus discriminantes impliquant des instances de la catégorie
   Philosophe $\rightarrow$ *affirmer, appeler, considérer, décrire, défendre, développer, fonder, publier, reprendre, écrire, etc.*

Profil relationnel



Profil relationnel



deFacto

1. extraction de tuples depuis 31M de phrases de Wikipedia-fr (2014) à l'aide d'une adaptation de ReVerb au français
2. arg_2s les plus fréquents de la relation *être* promus catégories
3. calcul du **profil relationnel** de chaque catégorie : les relations les plus discriminantes impliquant des instances de la catégorie
4. sélection des instances les plus représentatives de chaque catégorie à l'aide de leur profil
   Acteur $\rightarrow$ *David Arquette, Jeremy Bulloch, Fernand Gravey, etc.*

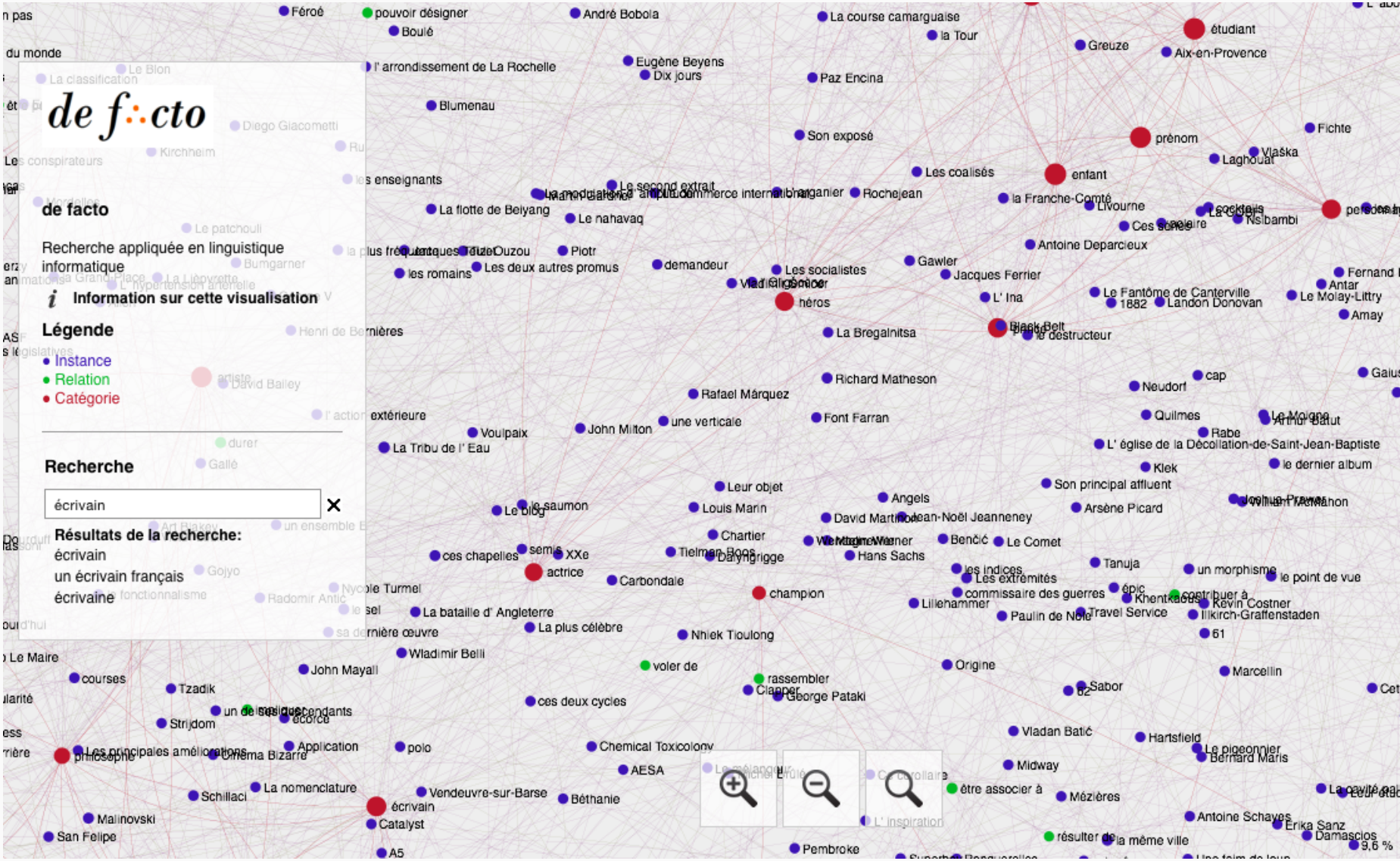
deFacto

1. extraction de tuples depuis 31M de phrases de Wikipedia-fr (2014) à l'aide d'une adaptation de ReVerb au français
2. arg_2s les plus fréquents de la relation *être* promus catégories
3. calcul du **profil relationnel** de chaque catégorie : les relations les plus discriminantes impliquant des instances de la catégorie
4. sélection des instances les plus représentatives de chaque catégorie à l'aide de leur profil
5. *clustering* (K-means) des catégories à l'aide de leur profil
   Écrivain → *artiste, auteur, créateur, dessinateur, historien, magazine, philosophe, élève*

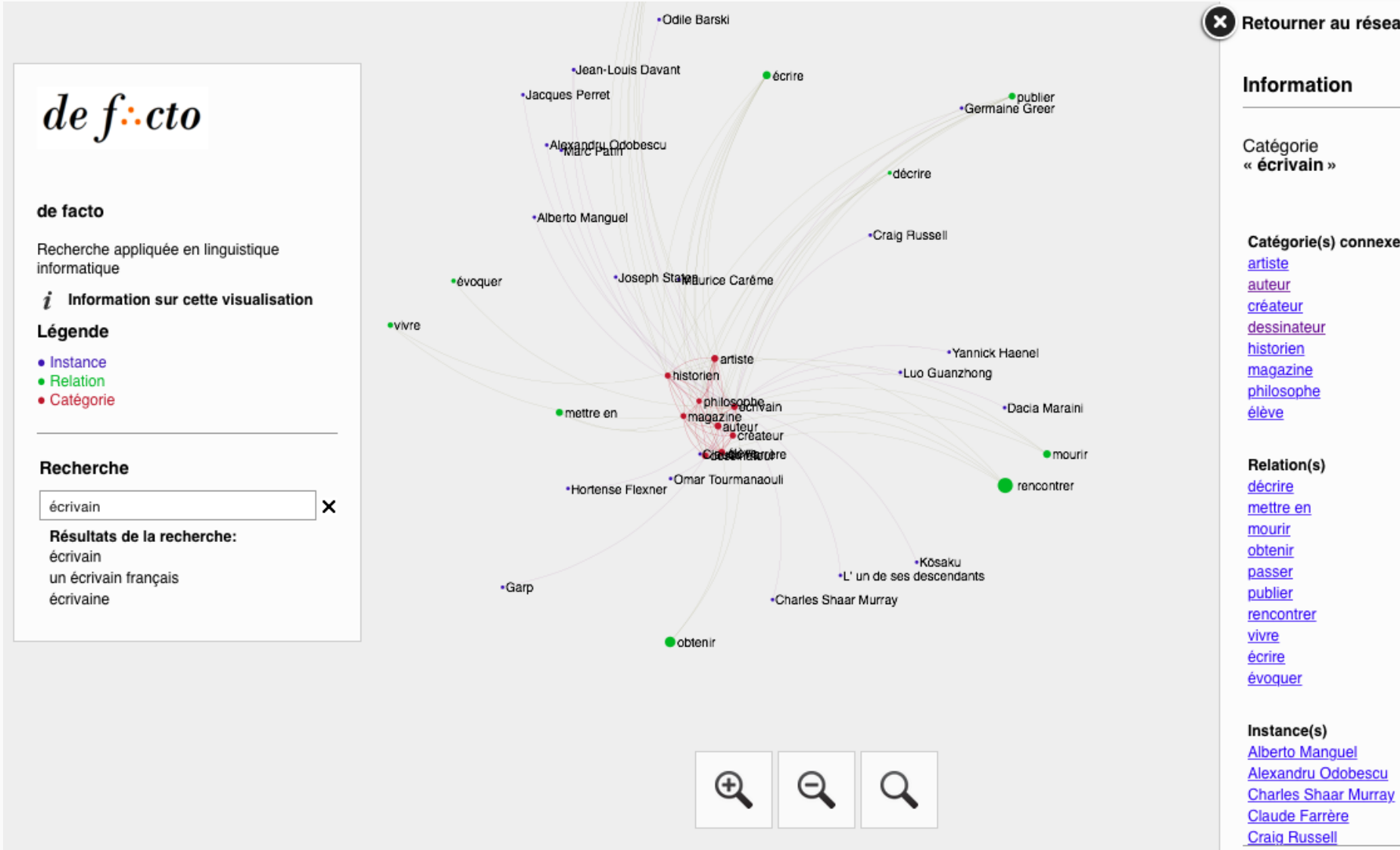
deFacto

1. extraction de tuples depuis 31M de phrases de WIKIPEDIA-FR (2014) à l'aide d'une adaptation de REVERB au français
2. arg_2s les plus fréquents de la relation *être* promus catégories
3. calcul du **profil relationnel** de chaque catégorie : les relations les plus discriminantes impliquant des instances de la catégorie
4. sélection des instances les plus représentatives de chaque catégorie à l'aide de leur profil
5. *clustering* (K-means) des catégories à l'aide de leur profil
6. génération d'un graphe navigable avec un *plugin* (`sigma.js`) modifié pour l'occasion

deFacto

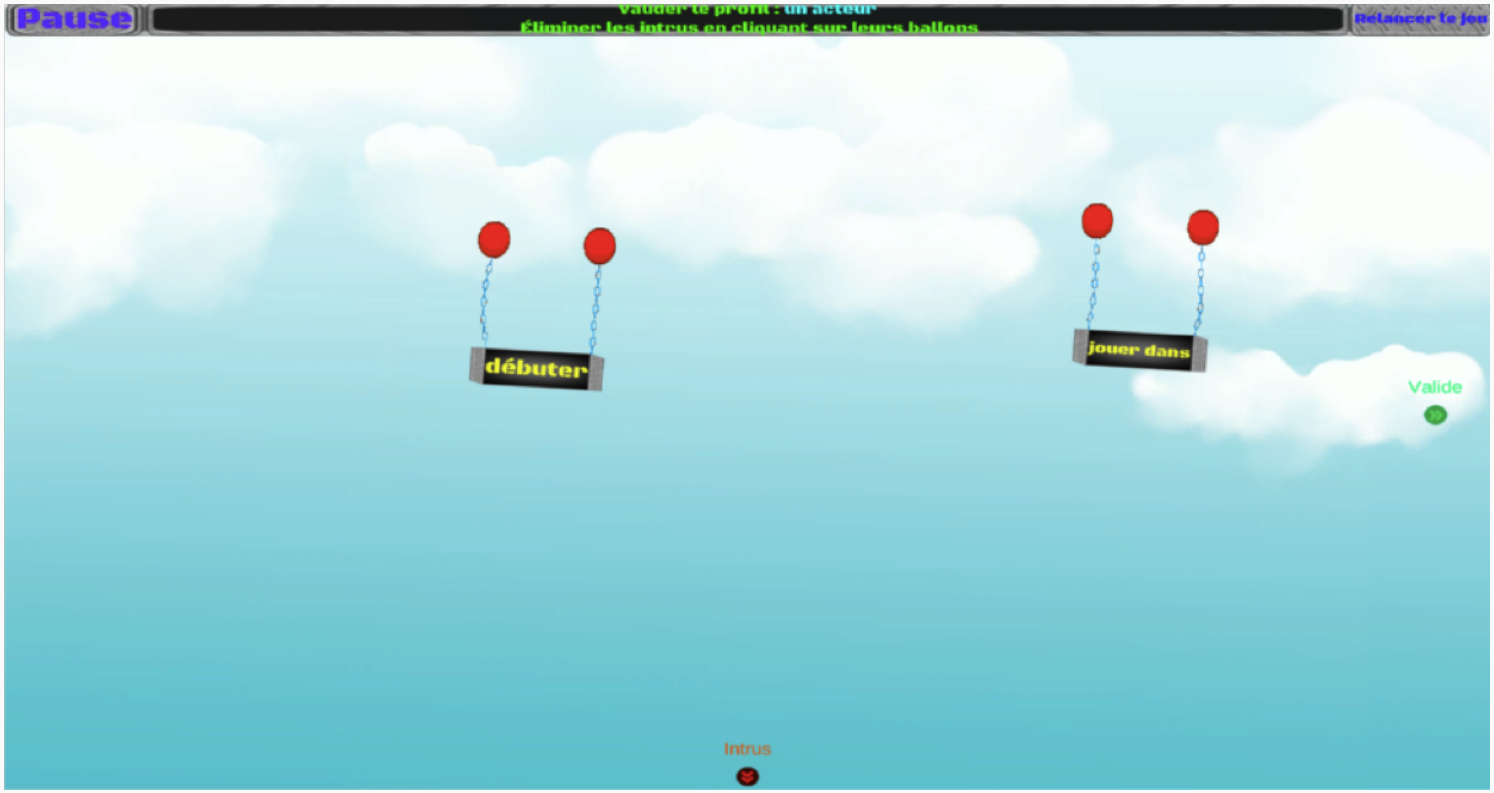

de f:cto
de facto
Recherche appliquée en linguistique informatique
Information sur cette visualisation
Légende
Instance
Relation
Catégorie
Recherche
écrivain
Résultats de la recherche:
écrivain
un écrivain français
écrivaine
Féroé
pouvoir désigner
André Bobola
La course camargualse
étudiant
Boulé
la Tour
Greuze
Aix-en-Provence
Eugène Beyens
Dix jours
Paz Encina
Blumenau
Son exposé
prénom
Fichte
Vlaška
Laghouat
Les coalisés
enfant
Rochejean
la Franche-Comté
Livourne
Ks bambi
La flotte de Belyang
Le nahavaq
Antoine Deparcieux
Piotr
Les deux autres promus
demandeur
Les socialistes
Gawler
Jacques Ferrier
héros
L' Ina
Le Fantôme de Canterville
1882
Landon Donovan
Le Molay-Littry
Antar
Amay
La Bregalnitsa
Richard Matheson
Rafael Márquez
Neudorf
cap
Quilmes
Font Farran
une verticale
John Milton
Voulpaix
La Tribu de l' Eau
Rabe
L' église de la Décollation-de-Saint-Jean-Baptiste
Klek
le dernier album
Son principal affluent
Leur objet
le saumon
Le blog
Louis Marin
Angels
Arsène Picard
David Martinon
Jean-Noël Jeanneney
Chartier
Benčić
Le Comet
ces chapelles
semis
XXe
actrice
Tielman Roos
Hans Sachs
Carbondale
champion
Tanuja
un morphisme
le point de vue
Les extrémités
commissaire des guerres
Lillehammer
Paulin de Noé
Travel Service
Ilkirch-Graffenstaden
61
La bataille d' Angleterre
La plus célèbre
Nhiek Tioulong
Wladimir Belli
voler de
Origine
Marcellin
courses
John Mayall
Tzadik
rassembler
George Pataki
Sabor
ces deux cycles
Strijdom
Vladan Batić
Hartsfield
Le pigeonnier
Bernard Maris
Chemical Toxicology
Application
polo
AESA
Midway
être associer à
Mézières
Schillaci
La nomenclature
Vendeuvre-sur-Barse
Béthanie
écrivain
Catalyst
Malinovski
San Felipe
A5
Pembroke
Antoine Schaves
Erika Sanz
Damascios
résulter de
la même ville

deFacto


de f::cto
de facto
Recherche appliquée en linguistique informatique
Information sur cette visualisation
Légende
Instance
Relation
Catégorie
Recherche
écrivain
Résultats de la recherche:
écrivain
un écrivain français
écrivaine
Odile Barski
Jean-Louis Davant
écrire
Jacques Perret
publier
Germaine Greer
Alexandru Odobescu
décrire
Alberto Manguel
Craig Russell
évoquer
Joseph Staline
Maurice Carême
vivre
artiste
historien
Yannick Haenel
Luo Guanzhong
philosophe
écrivain
mettre en
magazine
auteur
créateur
Dacia Maraini
mourir
Omar Tourmanaouli
Hortense Flexner
rencontrer
Kōsaku
L' un de ses descendants
Garp
Charles Shaar Murray
obtenir
Retourner au résea
Information
Catégorie
« écrivain »
Catégorie(s) connexe
artiste
auteur
créateur
dessinateur
historien
magazine
philosophe
élève
Relation(s)
décrire
mettre en
mourir
obtenir
passer
publier
rencontrer
vivre
écrire
évoquer
Instance(s)
Alberto Manguel
Alexandru Odobescu
Charles Shaar Murray
Claude Farrère
Craig Russell

deFacto : WikiGames [Forand and Langlais, 2017]

- 3 types de jeux pour aider à nettoyer la base de DEFACTO



érudit 20

Types de publications | Connexion | À propos | Recherche avancée | FR

Érudit célèbre ses 20 ans !

Recherche

Par auteur, titre, mots-clés...

→ **Revues**
→ **Livres et actes**
→ **Thèses et mémoires**
→ **Rapports de recherche**

Recherche avancée

Derniers numéros

Géographie

Les enjeux sociaux de l'eau : comparaisons

Philosophie, Théologie

LTP

Volume 74, numéro 1, février 2018

De la revue Laval théologique et philosophique

Anthropologie et ethnologie

Récits de savoirs partagés par l'art et la création en milieux autochtones

Géographie, Études urbaines

EUE
ENVIRONNEMENT URBAIN
URBAN ENVIRONMENT

La présence – absence des études urbaines en France
Volume 13, 2018

De la revue Environnement Urbain

Histoire

Numéro 180, mai–août 2018

Éducation

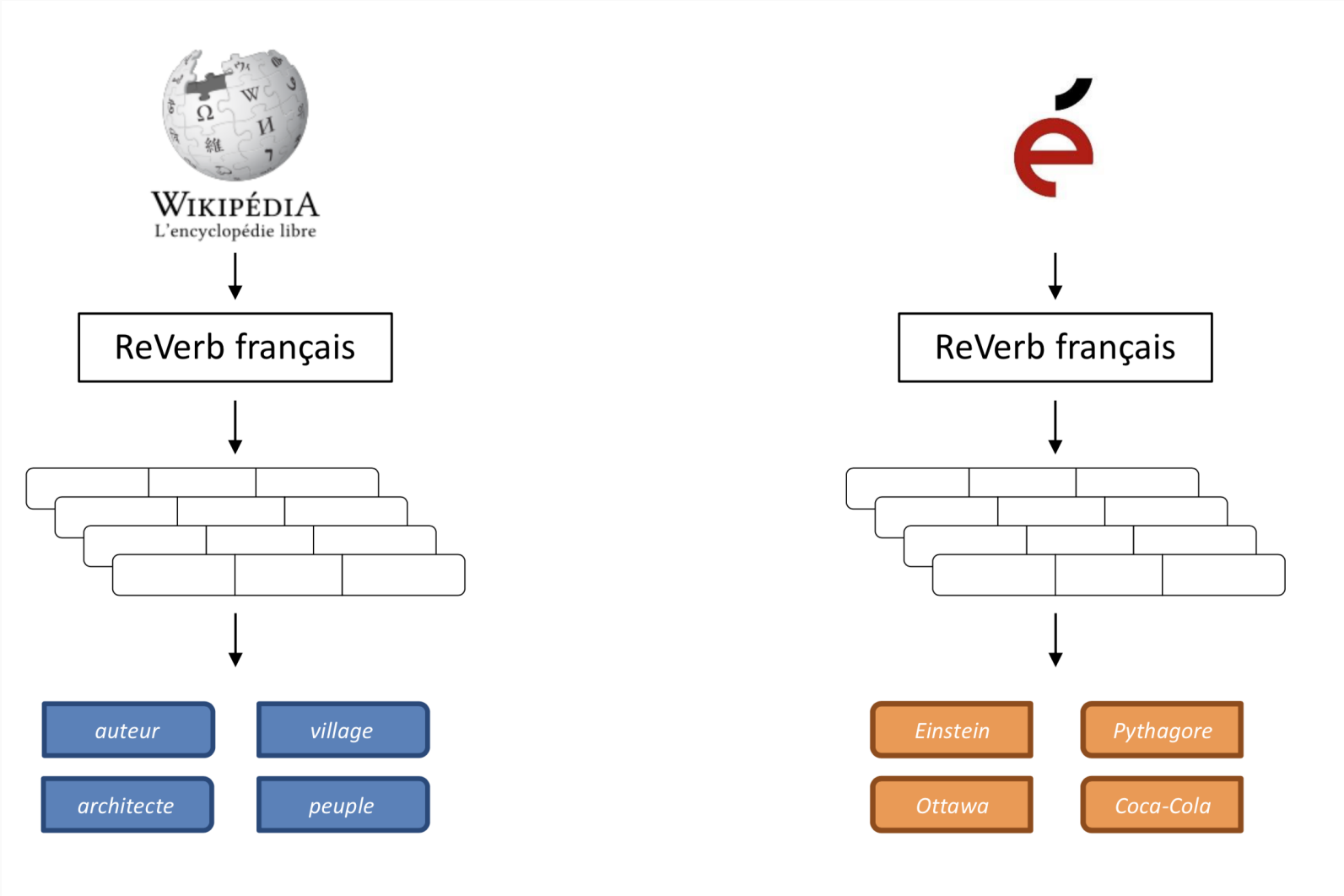
Rapport à l'écrit contextes de for (Vol. 2)
Volume 20, num 2017

De la revue Nou

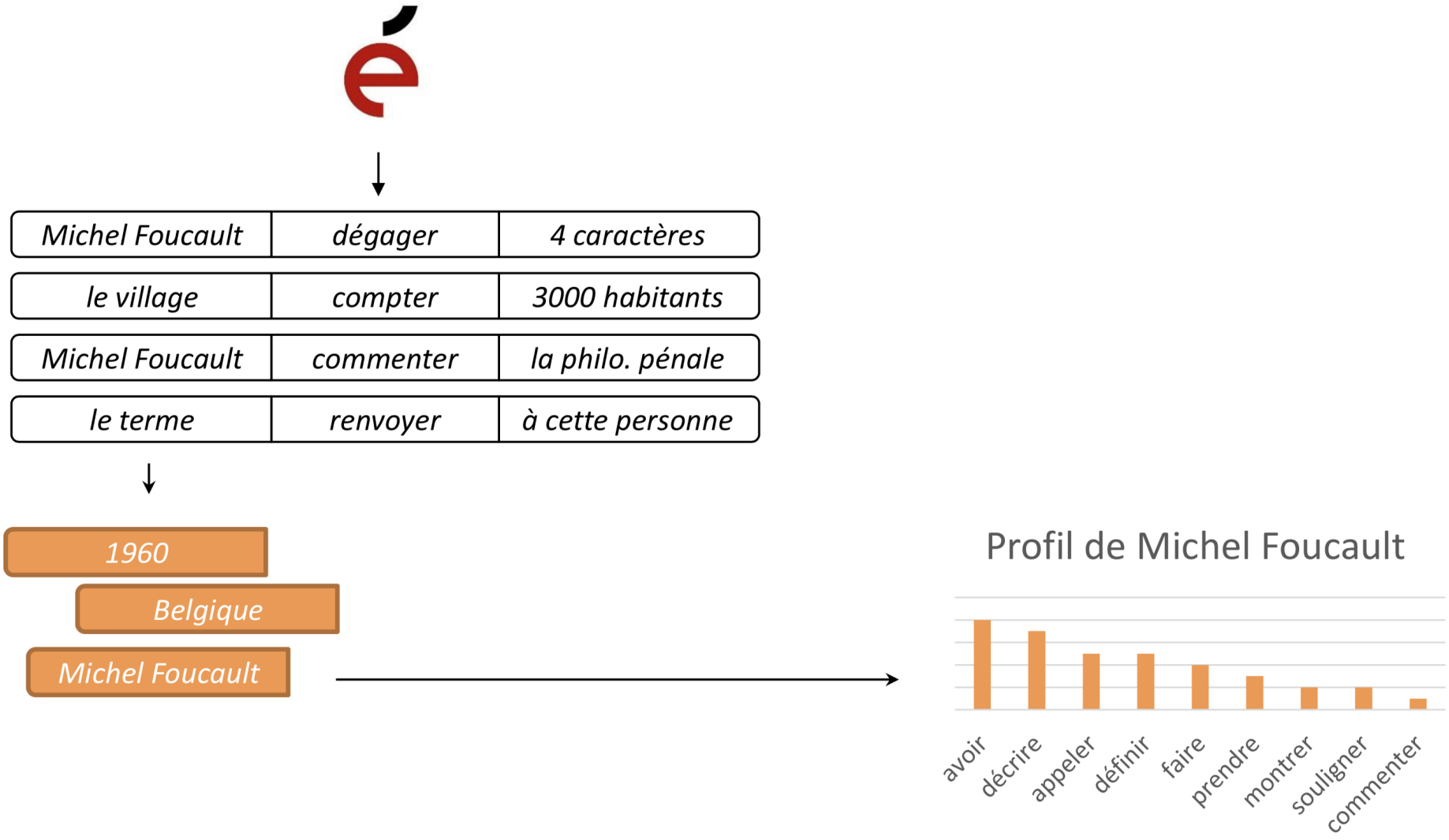
deFacto 2 Érudit [Gotti and Langlais, 2016]

- Peut-on ré-utiliser le typage émergeant de DEFACTO (WIKIPEDIA) pour étiqueter des entités extraites d'Érudit ?
  - 3M de phrases (français)
  - REVERB : 5M de triplets
  - TEST : 100 entités vues au moins 50 fois dans Érudit étiquetées à la main

deFacto 2 Érudit



deFacto 2 Érudit



deFacto 2 Érudit

70% de bon étiquetage

Instance	Type véritable	Type prédit
1960	année, période	période
Aragon	artiste, auteur, écrivain	auteur
article	étude, livre, recueil	étude
Belgique	pays, lieu, lieux, toponyme	ville
Canadiens français	peuple, groupe	personnages
Allemagne	pays, lieu, lieux, toponyme	ville
Santé Canada	organisation, association, organisme, groupe	médecin
Michel Foucault	philosophe, écrivain, auteur	auteur
Premier minister	ministre, président, maître	ministre
Socrate	philosophe, écrivain, auteur	ville
texte	étude, livre, recueil	livre

Intermède

Allium

Allium

Prototype **en cours** de développement réalisé dans le cadre d'un projet FCI piloté par Vincent Larivière (EBSI) dans le but d'améliorer la plateforme Érudit.

Allium

6 | EN COUVERTURE | ENTREVUE

SYLVAIN L'ESPÉRANCE

«LA GRÈCE, C'EST CE QUI NOUS ATTEND»

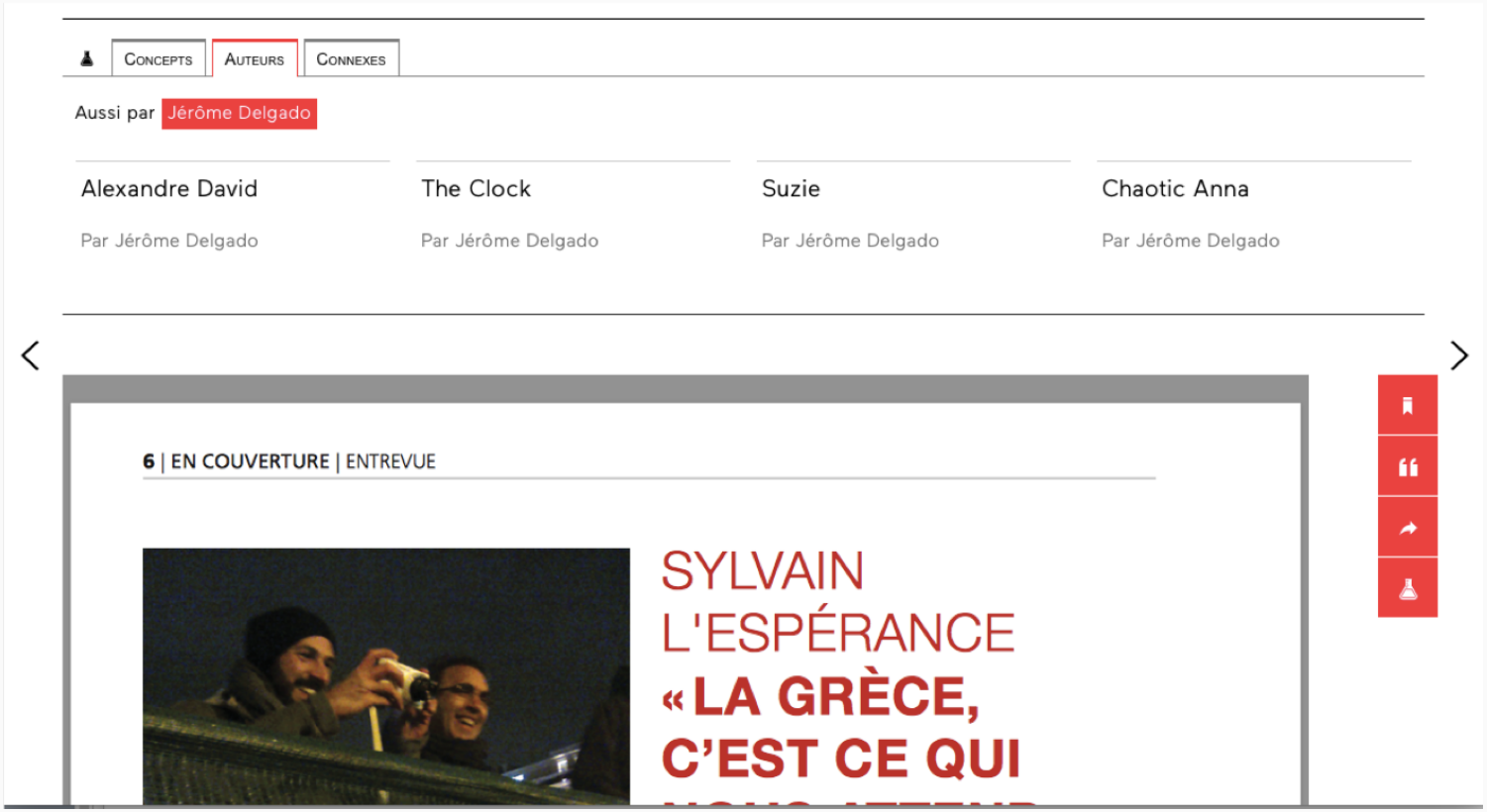
Dévoilé aux Rencontres internationales du documentaire de Montréal, en novembre 2016, ***Combat au bout de la nuit*** *est un film-fleuve porté par le désir de lutter contre les mesures d'austérité des gouvernements. Sylvain L'Espérance s'est tenu à le tourner en Grèce, mais il le considère comme un écho de l'état du monde. Sa présentation dans la section Panorama de la 67e Berlinale, en février 2017, le ravit particulièrement parce qu'on lui a annoncé avoir voulu faire, avec cette sélection, «un geste politique». Le documentariste québécois ne pouvait mieux être servi, lui qui cherche, derrière sa caméra imprégnée de poésie, à suggérer des formes de résistance.*

JÉRÔME DELGADO

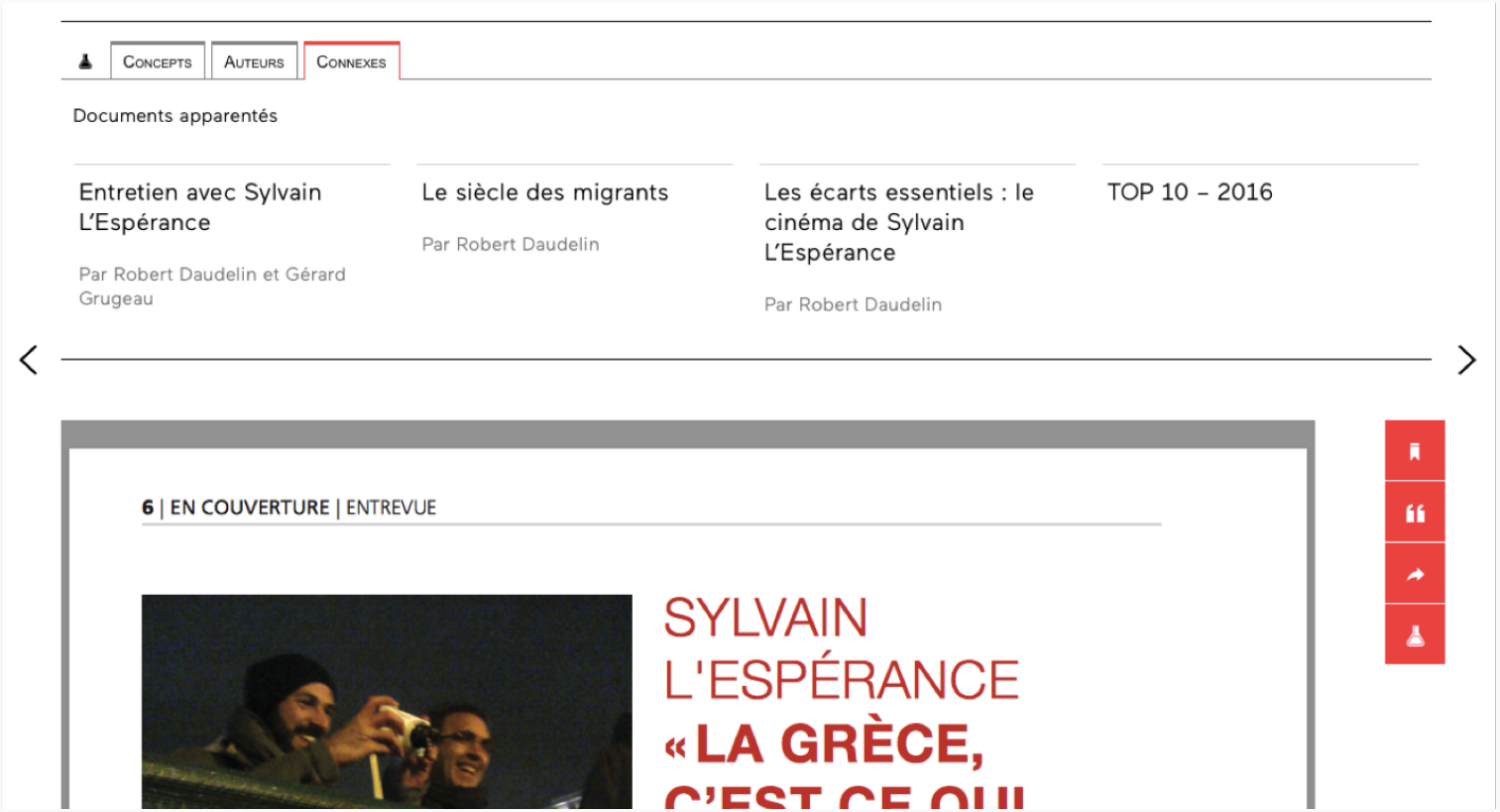
Allium


Numéro 307, Mars, 2017, p. 6–9
Combat au bout de la nuit
Tous droits réservés © La revue Séquences Inc., 2017
CONCEPTS
AUTEURS
CONNEXES
LIEUX
Grèce 43,7
Athènes 25,4
AUTRES CONCEPTS
Espérance 47,8
Combat 40,9
Sylvain L' Espérance 35,8
film 24,2
révolte 19,6
bout de la nuit 16,4

Allium


Concepts
Auteurs
Connexes
Aussi par Jérôme Delgado
Alexandre David
Par Jérôme Delgado
The Clock
Par Jérôme Delgado
Suzie
Par Jérôme Delgado
Chaotic Anna
Par Jérôme Delgado
6 | EN COUVERTURE | ENTREVUE
SYLVAIN
L'ESPÉRANCE
« LA GRÈCE,
C'EST CE QUI

Allium


Concepts
Auteurs
Connexes
Documents apparentés
Entretien avec Sylvain L'Espérance
Par Robert Daudelin et Gérard Grugeau
Le siècle des migrants
Par Robert Daudelin
Les écarts essentiels : le cinéma de Sylvain L'Espérance
Par Robert Daudelin
TOP 10 – 2016
6 | EN COUVERTURE | ENTREVUE
SYLVAIN L'ESPÉRANCE
«LA GRÈCE,

Structuration

Patty [Nakashole et al., 2012]

- Extracteur similaire à OLLIE où les arguments sont restreints aux seules entités nommées de YAGO
- tuples généralisés en patrons Syntaxic-Ontological-Lexical
  - séquence de mots, POS et $\star$
  - ⟨person⟩'s [adj] voice * ⟨song⟩ matche :
    Amy Winehouse's soft voice in 'Rehab'
    Elvis Presley's solid voice in this song 'All shook up'
- structuration selon la généralité des patrons, et leur synonymie
  $\langle person \rangle$ *winner of* $\langle award \rangle \Rightarrow \langle person \rangle$ *nominated for* $\langle award \rangle$

Patty

Relation	Paraphrases	Precision	Sample Paraphrases
DBPedia/artist	83	0.96±0.03	[adj] studio album of, [det] song by ...
DBPedia/associatedBand	386	0.74±0.11	joined band along, plays in ...
DBPedia/doctoralAdvisor	36	0.558±0.15	[det] student of, under * supervision ...
DBPedia/recordLabel	113	0.86±0.09	[adj] artist signed to, [adj] record label ...
DBPedia/riverMouth	31	0.83±0.12	drains into, [adj] tributary of ...
DBPedia/team	1,108	0.91±0.07	be * traded to, [prp] debut for ...
YAGO/actedIn	330	0.88±0.08	starred in * film, [adj] role for ...
YAGO/created	466	0.79±0.10	founded, 's book ...
YAGO/isLeaderOf	40	0.53±0.14	elected by, governor of ...
YAGO/holdsPoliticalPosition	72	0.73±0.10	[prp] tenure as, oath as ...

Table 6: Sample Results for Relation Paraphrasing

Webre [Min et al., 2012]

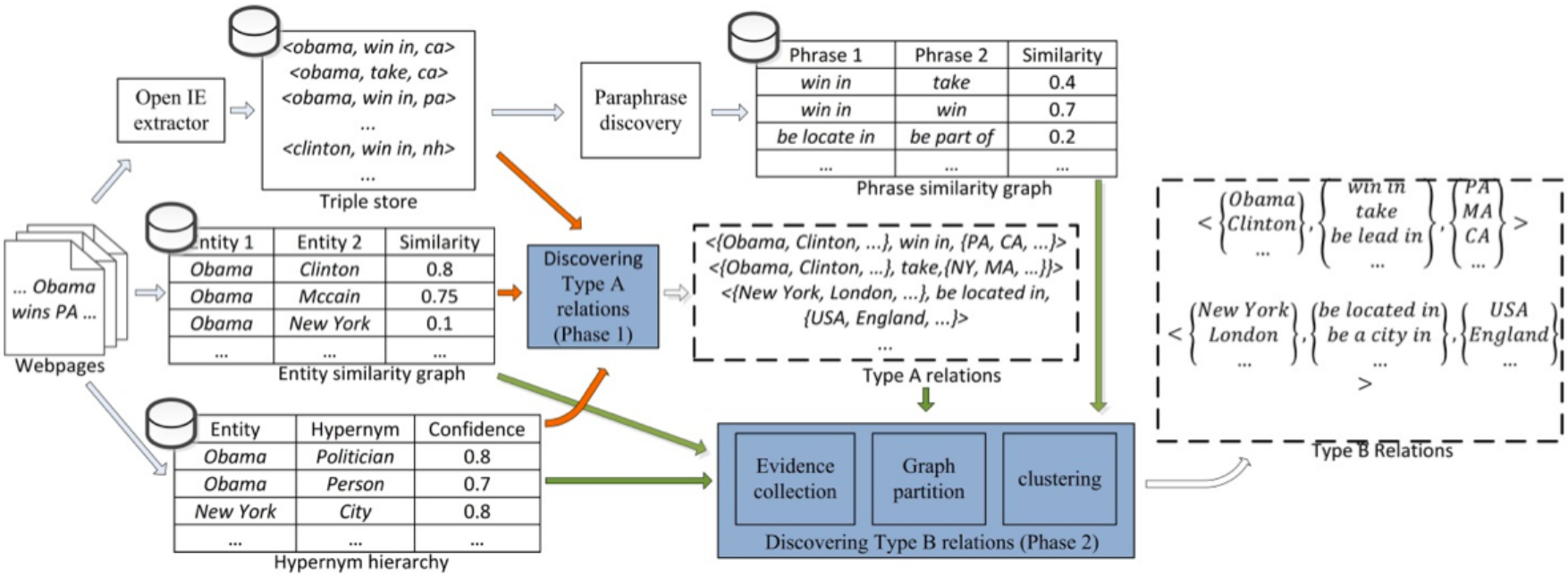


Figure 1. Overview of the WEBRE algorithm (Illustrated with examples sampled from experiment results). The tables and rectangles with a database sign show knowledge sources, shaded rectangles show the 2 phases, and the dotted shapes show the system output, a set of Type A relations and a set of Type B relations. The orange arrows denote resources used in phase 1 and the green arrows show the resources used in phase 2.

Webre [Min et al., 2012]

- 14.7M de triplets distincts extraits par ReVerb depuis ClueWeb (filtres appliqués)
  - 1.3M relations verbales, 3.3M entités (args)
- 0.2M relations de type A, 84 000 de type B

Argument 1	Relation phrase	Argument 2
marijuana, caffeine, nicotine...	*result in, be risk factor for, be major cause of...*	*insomnia, emphysema, breast cancer,...*
C# 2.0, php5, java, c++, ...	*allow the use of, also use, introduce the concept of...*	*destructors, interfaces, template,...*
clinton, obama, mccain, ...	*win, win in, take, be lead in,...*	*ca, dc, fl, nh, pa, va, ga, il, nc,...*

Table 3. Sample Type B relations extracted.

[Balasubramanian et al., 2013]

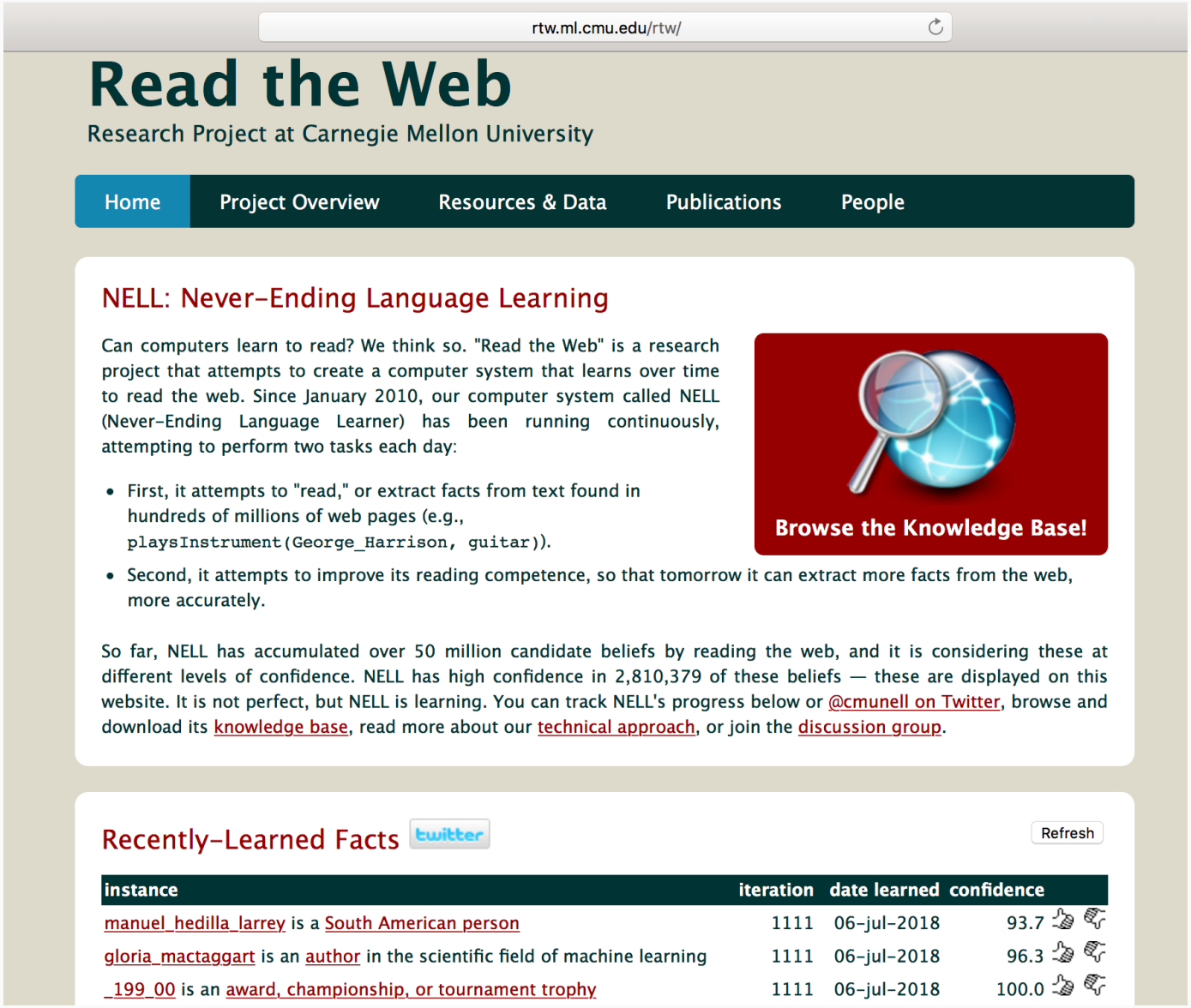
- Extraction de schémas d'événements à partir de textes

Actor	Rel	Actor
A1:<person>	failed	A2:test
A1:<person>	was suspended for	A3:<time period>
A1:<person>	used	A4:<substance, drug>
A1:<person>	was suspended for	A5:<game, activity>
A1:<person>	was in	A6:<location>
A1:<person>	was suspended by	A7:<org, person>
Actor Instances:		
A1: {Murray, Morgan, Governor Bush, Martin, Nelson}		
A2: {test}		
A3: {season, year, week, month, night}		
A4: {cocaine, drug, gasoline, vodka, sedative}		
A5: {violation, game, abuse, misfeasance, riding}		
A6: {desert, Simsbury, Albany, Damascus, Akron}		
A7: {Fitch, NBA, Bud Selig, NFL, Gov Jeb Bush}		

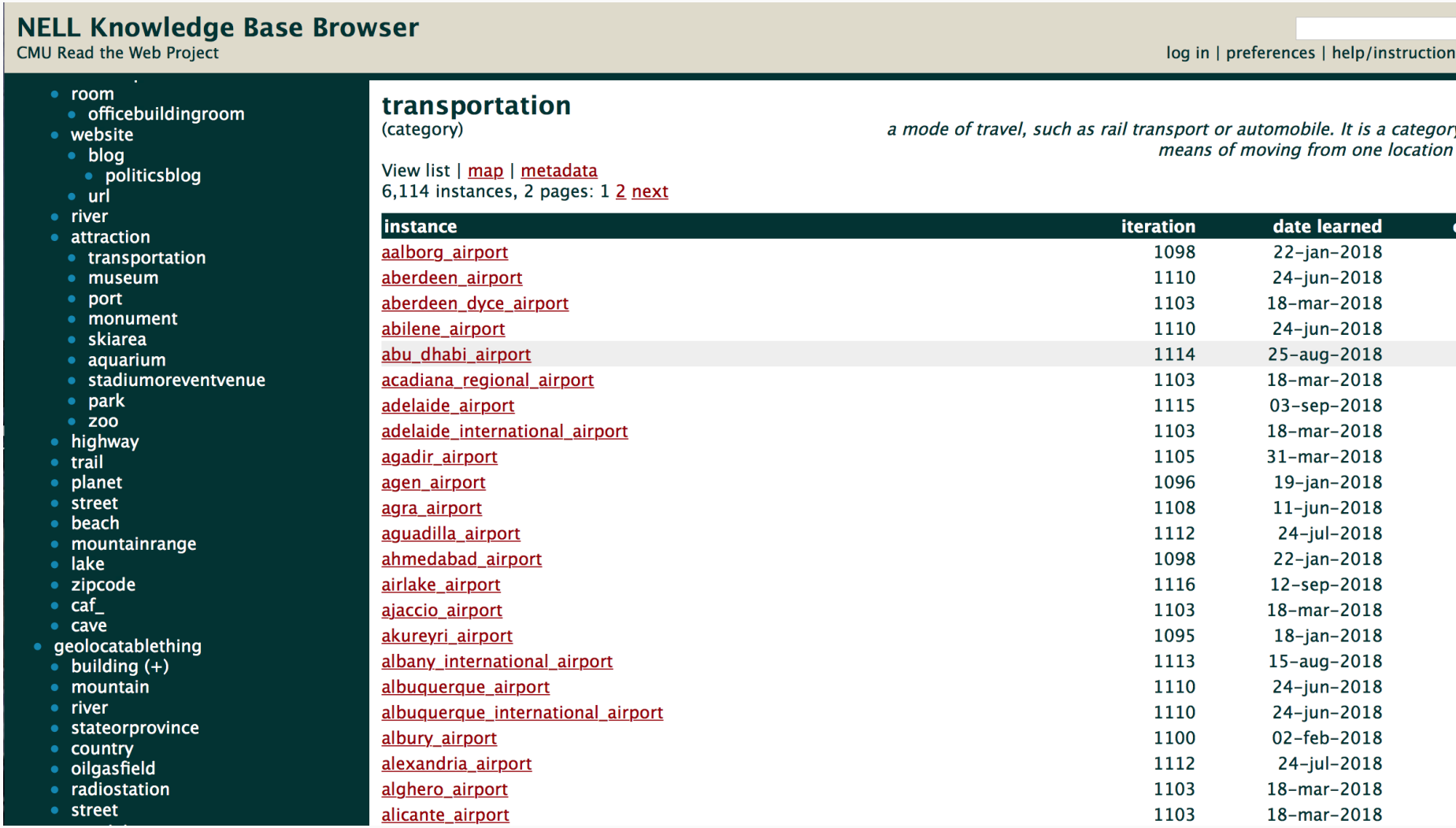
Table 1: An event schema produced by our system, represented as a set of $(Actor, Rel, Actor)$ triples, and a set of instances for each actor $A1$, $A2$, etc. For clarity we show unstemmed verbs.

Applications

Never Ending Language Learning (NELL)



Never Ending Language Learning (NELL)



NELL Knowledge Base Browser
CMU Read the Web Project

log in | preferences | help/instruction

- room
  - officebuildingroom
- website
  - blog
    - politicsblog
  - url
- river
- attraction
  - transportation
  - museum
  - port
  - monument
  - skiarea
  - aquarium
  - stadiumoreventvenue
  - park
  - zoo
- highway
- trail
- planet
- street
- beach
- mountainrange
- lake
- zipcode
- caf_
- cave
- geolocatablething
  - building (+)
  - mountain
  - river
  - stateorprovince
  - country
  - oilgasfield
  - radiostation
  - street

transportation

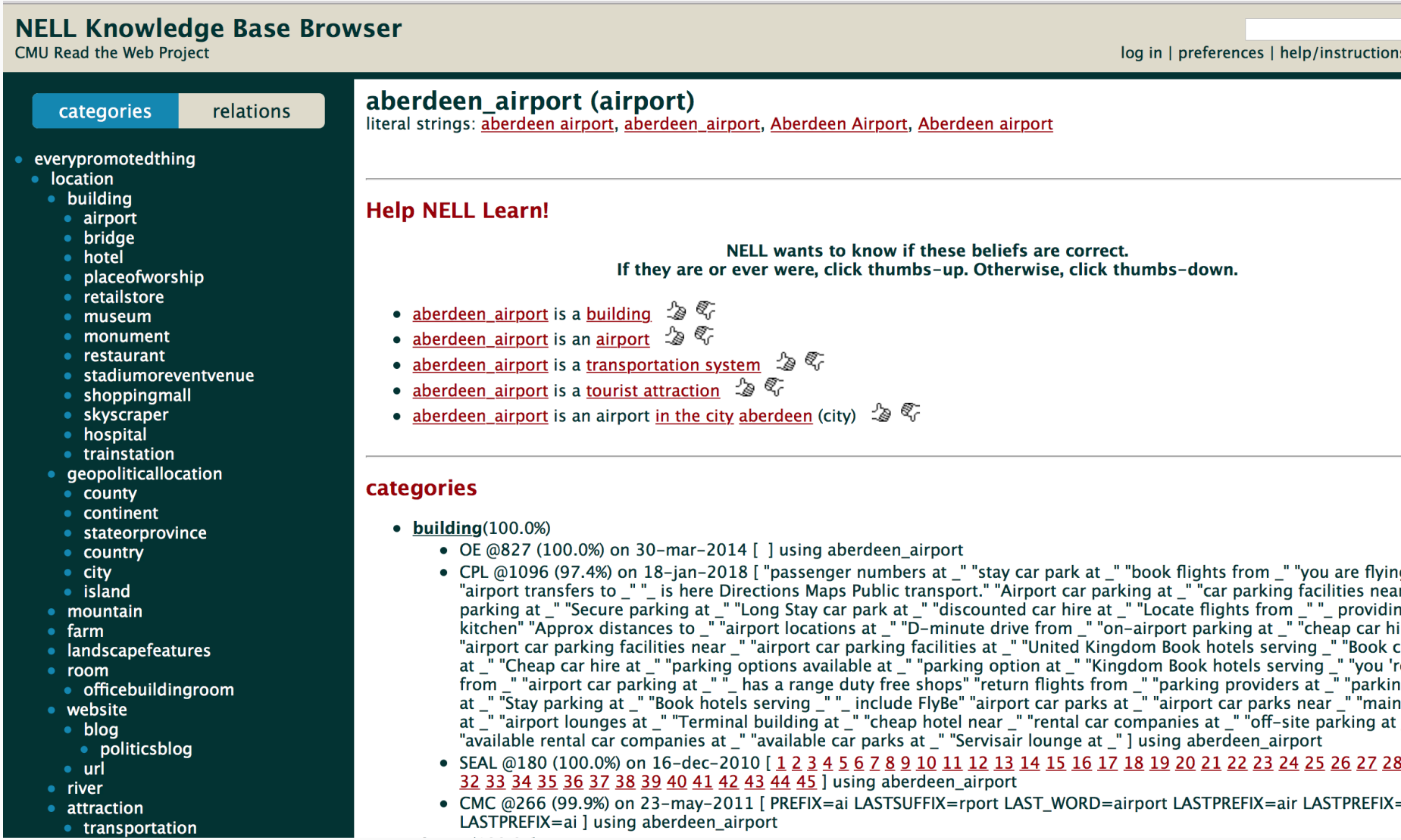
(category)

a mode of travel, such as rail transport or automobile. It is a categor
means of moving from one location

View list | map | metadata
6,114 instances, 2 pages: 1 2 next

instance	iteration	date learned
aalborg_airport	1098	22-jan-2018
aberdeen_airport	1110	24-jun-2018
aberdeen_dyce_airport	1103	18-mar-2018
abilene_airport	1110	24-jun-2018
abu_dhabi_airport	1114	25-aug-2018
acadiana_regional_airport	1103	18-mar-2018
adelaide_airport	1115	03-sep-2018
adelaide_international_airport	1103	18-mar-2018
agadir_airport	1105	31-mar-2018
agen_airport	1096	19-jan-2018
agra_airport	1108	11-jun-2018
aguadilla_airport	1112	24-jul-2018
ahmedabad_airport	1098	22-jan-2018
airlake_airport	1116	12-sep-2018
ajaccio_airport	1103	18-mar-2018
akureyri_airport	1095	18-jan-2018
albany_international_airport	1113	15-aug-2018
albuquerque_airport	1110	24-jun-2018
albuquerque_international_airport	1110	24-jun-2018
albury_airport	1100	02-feb-2018
alexandria_airport	1112	24-jul-2018
alghero_airport	1103	18-mar-2018
alicante_airport	1103	18-mar-2018

Never Ending Language Learning (NELL)

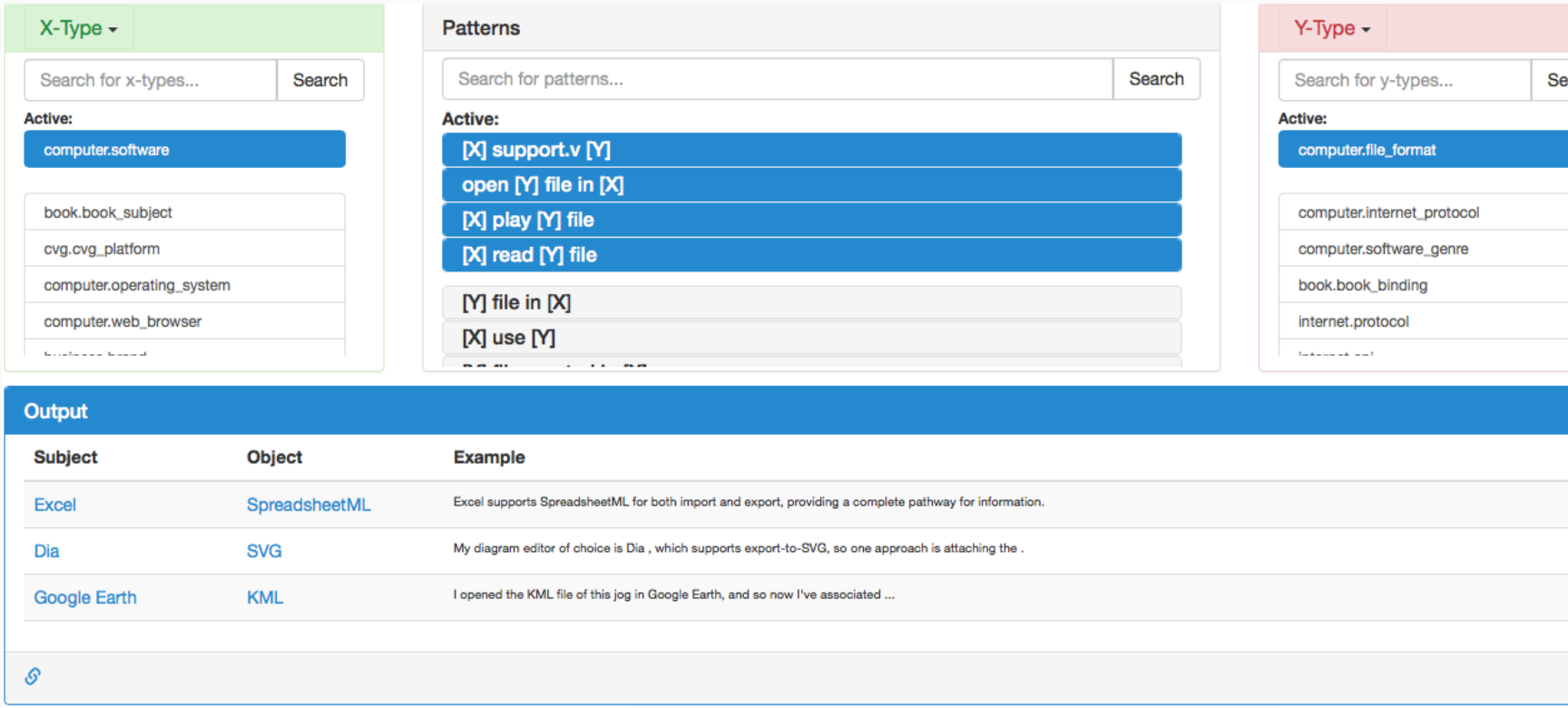


Never Ending Language Learning (NELL)

So far, NELL has accumulated over 50 million candidate beliefs by reading the web, and it is considering these at different levels of confidence. NELL has high confidence in 2,810,379 of these beliefs — these are displayed on this website. It is not perfect, but NELL is learning. You can track NELL's progress below or @cmunell on Twitter, browse and download its knowledge base, read more about our technical approach, or join the discussion group.

Exploration de grandes bases de textes [Akbik et al., 2014]

- basé sur les types de Yago
- aide au développement d'extracteurs



Literome project [Poon et al., 2014]

- extraction of genomic knowledge from PubMed articles
- available in the cloud for browsing, searching and reasoning

Open-domain QA [Fader et al., 2014]

- QA $\equiv$ séquence d'opérateurs qui transforment une question en une requête à un *tuple-store*
  - 4 $\neq$ tuple-stores : FreeBase et 3 autres extraits automatiquement
- une fonction donnant un score à chaque opération est apprise sur une corpus de questions et leur réponse.
- répondre $\equiv$ chercher
  dans l'espace des opérations

Open-domain QA [Fader et al., 2014]

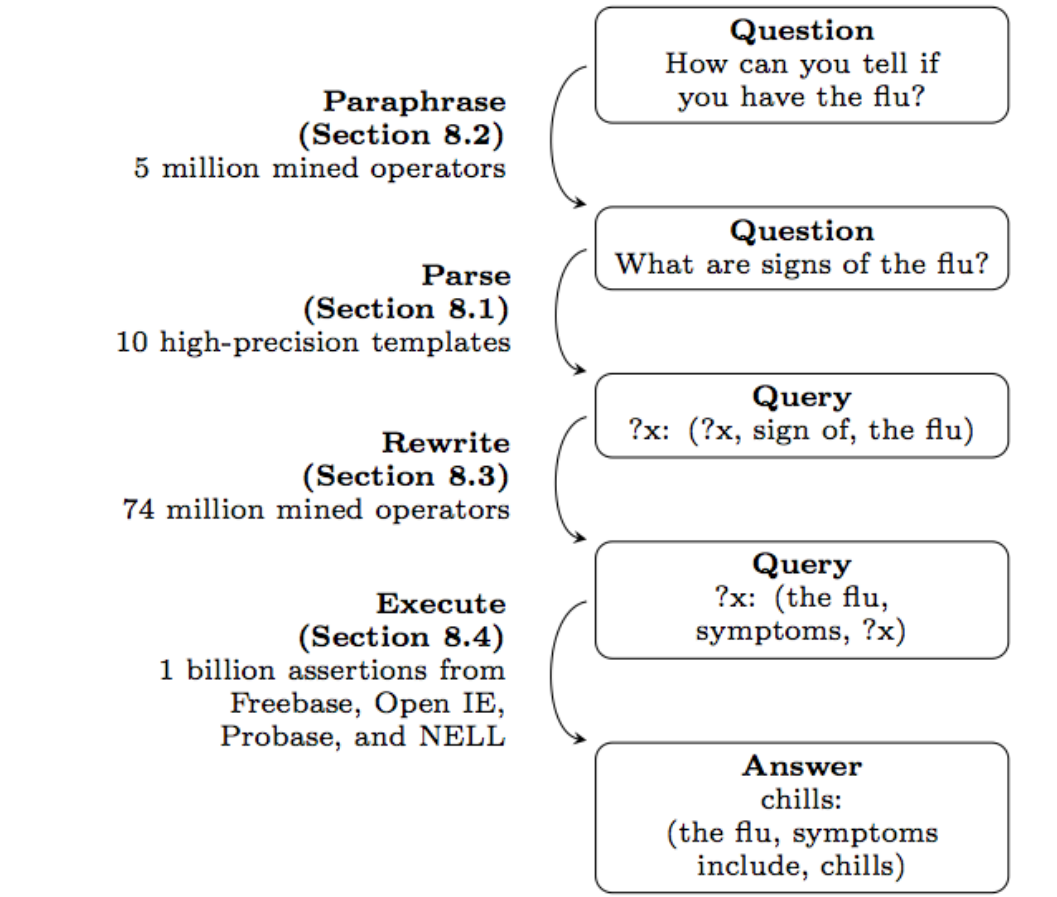


Figure 1: OQA automatically mines millions of operators (left) from unlabeled data, then learns to compose them to answer questions (right) using evidence from multiple knowledge bases.

Open-domain QA [Fader et al., 2014]

Input	What are some examples of building maintenance jobs ?
Parse	`?x: (?x, example of, building maintenance jobs)`
Rewrite	`?x: (?x, is-a, building maintenance job)`
Execute	{changing light bulb, is-a, small building maintenance job}

Input	What animal represents California ?
Paraphrase	What are California's symbols ?
Parse	`?x: (california,symbols, ?x)`
Execute	{California Water Service, Trading symbol, CWT}

Conclusion

Conclusion

- Extraire et exploiter des connaissances extraites de textes est un **enjeu** académique et industriel

Conclusion

- Extraire et exploiter des connaissances extraites de textes est un **enjeu** académique et industriel
- Nombreuses ressources disponibles :
  - Yago Freebase, ConcepNet, etc.
  - s'y lier est important (*entity linking*)

Conclusion

- Extraire et exploiter des connaissances extraites de textes est un **enjeu** académique et industriel
- Nombreuses ressources disponibles :
  - Yago Freebase, ConcepNet, etc.
  - s'y lier est important (*entity linking*)
- Évaluer la technologie d'extraction est difficile
  - arbitraire, **rappel** difficile à mesurer

Conclusion

- Extraire et exploiter des connaissances extraites de textes est un **enjeu** académique et industriel
- Nombreuses ressources disponibles :
  - Yago Freebase, ConcepNet, etc.
  - s'y lier est important (*entity linking*)
- Évaluer la technologie d'extraction est difficile
  - arbitraire, **rappel** difficile à mesurer
- pas ou peu de normalisation des relations extraites

Conclusion

- Extraire et exploiter des connaissances extraites de textes est un **enjeu** académique et industriel
- Nombreuses ressources disponibles :
  - Yago Freebase, ConcepNet, etc.
  - s'y lier est important (*entity linking*)
- Évaluer la technologie d'extraction est difficile
  - arbitraire, **rappel** difficile à mesurer
- pas ou peu de normalisation des relations extraites
- Beaucoup d'information textuelle difficile d'accès
  - importance des informations temporelles

Chilly Gonzales

"Chilly Gonzales (born Jason Charles Beck ; 20 March 1972) is a Grammy-winning Canadian musician who resided in Paris, France for several years, and now lives in Cologne, Germany."

Chilly Gonzales

"Chilly Gonzales (born Jason Charles Beck ; 20 March 1972) is a Grammy-winning Canadian musician who resided in Paris, France for several years, and now lives in Cologne, Germany."

⟨Chilly Gonzales, [was] born, Jason Charles Beck⟩
⟨Chilly Gonzales, [was] born [on], 20 March 1972⟩
⟨Chilly Gonzales, is a, musician⟩
⟨Chilly Gonzales, [won], Grammy [award]⟩
⟨Chilly Gonzales, is, Canadian⟩
⟨Chilly Gonzales, resided in, Paris⟩ **for several years**
⟨Paris, [is in], France⟩
⟨Chilly Gonzales, lives in, Cologne⟩
⟨Cologne, [is in], Germany⟩


Agichtein, E. and Gravano, L. (2000).
Snowball : Extracting relations from large plain-text collections.
In *Fifth ACM Conference on Digital Libraries*, pages 85–94.


Akbik, A., Michael, T., and Boden, C. (2014).
Exploratory relation extraction in large text corpora.
In *25th International Conference on Computational Linguistics*, pages 2087–2096.


Angeli, G., Johnson Premkumar, M. J., and Manning, C. D. (2015).
Leveraging linguistic structure for open domain information extraction.
In *53rd Annual Meeting of the Association for Computational Linguistics and the 7th International Joint Conference on*

Natural Language Processing (Volume 1 : Long Papers), pages 344–354.

 Balasubramanian, N., Soderland, S., Mausam, and Etzioni, O. (2013).
Generating coherent event schemas at scale.
In *EMNLP'13*, pages 1721–1731.

 Banko, M., Cafarella, M. J., Soderland, S., Broadhead, M., and Etzioni, O. (2007).
Open information extraction from the web.
In *IN IJCAI*, pages 2670–2676.

 Clark, P., Cowhey, I., Etzioni, O., Khot, T., Sabharwal, A., Schoenick, C., and Tafjord, O. (2018).
Think you have solved question answering ? try arc, the AI2 reasoning challenge.
CoRR.


Fader, A., Soderland, S., and Etzioni, O. (2011).
Identifying relations for open information extraction.
In *Empirical Methods in Natural Language Processing*, EMNLP '11, pages 1535–1545.


Fader, A., Zettlemoyer, L., and Etzioni, O. (2014).
Open question answering over curated and extracted knowledge bases.
In *Proceedings of the 20th ACM SIGKDD International Conference on Knowledge Discovery and Data Mining*, KDD '14, pages 1156–1165.


Forand, K. and Langlais, P. (2017).
Curating an open information extraction knowledge base using games with a purpose.
In *Game4NLP, workshop at EACL 2017*, Valencia, Spain.


Gotti, F. and Langlais, P. (2016).
From french wikipedia to erudit : A test case for cross-domain open information extraction.
Computational Intelligence.


GuoDong, Z., Jian, S., Jie, Z., and Min, Z. (2005).
Exploring various knowledge in relation extraction.
In *43rd Annual Meeting on Association for Computational Linguistics*, pages 427–434.


Hearst, M. A. (1992).
Automatic acquisition of hyponyms from large text corpora.
In *Proceedings of the 14th Conference on Computational Linguistics - Volume 2*, COLING '92, pages 539–545.


Lai, G., Xie, Q., Liu, H., Yang, Y., and Hovy, E. (2017).

Race : Large-scale reading comprehension dataset from examinations.

In *Conference on Empirical Methods in Natural Language Processing*, pages 785–794. Association for Computational Linguistics.

Mausam, Schmitz, M., Bart, R., Soderland, S., and Etzioni, O. (2012).

Open language learning for information extraction.

In *Joint EMNLP and CoNLL*, pages 523–534.

Min, B., Shi, S., Grishman, R., and Lin, C.-Y. (2012).

Ensemble semantics for large-scale unsupervised relation extraction.

In *Proceedings of the 2012 Joint EMNLP and CoNLL*, pages 1027–1037.

Nakashole, N., Weikum, G., and Suchanek, F. (2012).

Patty : A taxonomy of relational patterns with semantic types.

In *Joint EMNLP and CoNLL*, pages 1135–1145.

Niklaus, C., Cetto, M., Freitas, A., and Handschuh, S. (2018).

A survey on open information extraction.

In *27th International Conference on Computational Linguistics*, pages 3866–3878.

Poon, H., Quirk, C., DeZiel, C., and Heckerman, D. (2014).

Literome : Pubmed-scale genomic knowledge base in the cloud.

Bioinformatics, 30(19) :2840–2842.

Wu, F. and Weld, D. S. (2010).

Open information extraction using wikipedia.

In *48th Annual Meeting of the Association for Computational Linguistics*, pages 118–127.