

---

# Travail Pratique 1

Par : Étienne Dubé et Alexis Miron

---

## Objectifs

Nous voulons concevoir un programme capable de détecter automatiquement les noms de lieux présents dans un texte.

## Échantillons

Nous travaillons sur l'extrait d'un texte issu de Wikipédia portant sur l'intelligence artificielle<sup>1</sup>. Le texte totalise à un peu plus de 4600 mots. Deux extraits du texte sont disponibles en annexe I.

Nous nous sommes basés sur la valeur sémantique de chaque mot afin de définir s'il s'agissait d'un lieu. Nous voulions seulement sélectionner les mots ayant comme référent un lieu physique réel et non une institution ou un lieu fictif.

## Observations

Nous avons retenus les mots suivants comme noms d'un lieu physique réel pouvant être placé sur une carte<sup>2</sup> :

NB : Le symbole “□” dénote un début/fin de phrase.

marqué par une conférence au	Dartmouth College	à l'été 1956 à
recherche se fait principalement aux	États-Unis	, notamment à l'université Stanford
aux États-Unis, notamment à l'	université Stanford	sous l'impulsion de John
impulsion de John McCarthy au	MIT	sous celle de Marvin Minsky
de Marvin Minsky, à l'	université Carnegie-Mellon	sous celle de Allen Newell
Herbert Simon et à l'	université d'Édimbourg	sous celle de Donald Michie
celle de Donald Michie, en	Europe	et en Chine, ainsi qu'
Michie en Europe et en	Chine	, ainsi qu'au Japon avec
en Chine, ainsi qu'au	Japon	avec le projet « ordinateurs de
□ En	France	l'un des pionniers est
800 avant notre ère, en	Égypte	. □

---

<sup>1</sup> [https://fr.wikipedia.org/wiki/Intelligence\\_artificielle](https://fr.wikipedia.org/wiki/Intelligence_artificielle)

<sup>2</sup> Il peut y avoir plus d'une occurrence d'un même mot.

de bronze Talos, gardien des	rivages de la Crète	, était parfois considéré comme une
ingénieurs fondée par Ctesibius à	Alexandrie	, et concevant des mécanismes destinés
exposé pendant plusieurs années en	France	, en Italie et en Angleterre
plusieurs années en France, en	Italie	et en Angleterre, et la
France, en Italie et en	Angleterre	, et la transparence de l
Malheureusement, l'incendie du	musée de Nijni Novgorod	en Russie vers 1879 détruit
musée de Nijni Novgorod en	Russie	vers 1879 détruit cet automate.
En juillet 2009, en	Californie	une conférence organisée par l
permanente dans le domaine à	Paris	. □
et de compréhension de l'	université Stanford	. □
□ En	France	□
□ En	France	, les pionniers sont Alain Colmerauer
est créé en 1979 à	Toulouse	. □
Conference on Artificial Intelligence à	Chambéry	en 1993, et la création
la République Emmanuel Macron au	Collège de France	pour annoncer la stratégie de

Nous avons refusés certains mots qui, de prime abord, pourraient sembler être des noms de lieux :

Contexte	Justification
[...] des objets jusqu'à l' <u>Olympe</u> et de revenir seuls [...]	"l' <i>Olympe</i> " fait référence au lieu fictif de la mythologie grec et non à la montagne existant en Grèce.
[...] pour amuser le roi <u>de France</u> , François I. □	" <i>de France</i> " est utilisé en tant que partie d'un titre royal, pour désigner les sujets du roi, et ne fait donc pas directement référence au pays.
□ Gio Battista Aleotti et Salomon <u>de Caus</u> , eux, ont construit [...]	" <i>de Caus</i> " est utilisé comme patronyme de l'homme en question, sans faire référence directement à sa ville d'origine.
□ En 2004, l' <u>Institut Singularity</u> a lancé une campagne [...]	"l' <i>Institut Singularity</i> " fait référence à l'institution et non au lieu

□ En 2009, le <u>Massachusetts Institute of Technology</u> (MIT) a lancé un projet [...]	“ <i>l’Institut Singularity</i> ” fait référence à l’institution et non au lieu
[...] recherches dans les <u>locaux de la NASA</u> . □	“ <i>les locaux de la NASA</i> ” fait référence à un type de lieu, dans un des plusieurs locaux que possède la NASA
[...] une audition se tient au <u>Sénat</u> : □	“ <i>au Sénat</i> ” fait référence à l’institution
[...] grâce au projet « <u>France IA</u> ». □	“ <i>France IA</i> ” fait référence à un mouvement
[...] annoncer la stratégie de <u>la France</u> dans ce domaine. □	“ <i>la France</i> ” fait référence au gouvernement de ce pays
□ En mars 2018, <u>Microsoft France</u> lance l’École IA Microsoft, [...]	“ <i>Microsoft France</i> ” fait référence à une branche d’une compagnie

---

Première itération

---

## Analyse

Suite à l’analyse des mots cibles et refusés durant la phase d’observation, nous avons fait les constats suivant :

- Les noms de lieux comportent un mot “central” avec une première lettre en majuscule.
- Le mot central est souvent précédé d’une préposition locative (“en”, “au”, “aux”, “à”)
- Certains nombres romains correspondent aux deux critères ci-dessus.

## Algorithme et Script Python

- 1) Stocker le texte dans la variable *texte*.
- 2) Couper le texte en mots distinct sur la base des espaces et de la ponctuation.
- 3) Définir les prépositions locatives introduisant typiquement des lieux (*en, au, aux, à*).
- 4) Itérer mot à mot sur le texte.
- 5) Pour chaque mot commençant par une majuscule, vérifier si le mot précédent est une des prépositions.
- 6) Imprimer les mots retenus.

```
import re

texte = open('IApourTP1.txt',encoding='utf8').read()
mots = re.split(r'[^\\w|-]+' ,texte)
```

```
prep_avant = ["en", "au", "aux", "à"]
```

```
for (idx, mot) in enumerate (mots) :
    if (len(mot) > 0 and mot[0].isupper()) :
        if (mots[idx-1].lower() in prep_avant) :
            print(mots[idx-5:idx+6])
```

## Évaluation

**Vert** → Mot cible retenu correctement

**Jaune** → Mot cible non retenu ou seulement en partie

**Rouge** → Mot retenu incorrectement

marqué par une conférence au	Dartmouth College	à l'été 1956 à
recherche se fait principalement aux	États-Unis	, notamment à l'université Stanford
aux États-Unis, notamment à l'	université Stanford	sous l'impulsion de John
impulsion de John McCarthy au	MIT	sous celle de Marvin Minsky
de Marvin Minsky, à l'	université Carnegie-Mellon	sous celle de Allen Newell
Herbert Simon et à l'	université d'Édimbourg	sous celle de Donald Michie
celle de Donald Michie, en	Europe	et en Chine, ainsi qu'
Michie en Europe et en	Chine	, ainsi qu'au Japon avec
en Chine, ainsi qu'au	Japon	avec le projet « ordinateurs de
□ En	France	l'un des pionniers est
800 avant notre ère, en	Égypte	. □
de bronze Talos, gardien des	rivages de la Crète	, était parfois considéré comme une
ingénieurs fondée par Ctesibius à	Alexandrie	, et concevant des mécanismes destinés
□ On a prêté à	Roger	Bacon la conception d'automates
exposé pendant plusieurs années en	France	, en Italie et en Angleterre
plusieurs années en France, en	Italie	et en Angleterre, et la
France, en Italie et en	Angleterre	, et la transparence de l
Malheureusement, l'incendie du	musée de Nijni Novgorod	en Russie vers 1879 détruisit
musée de Nijni Novgorod en	Russie	vers 1879 détruisit cet automate.

□ Gottfried Wilhelm Leibniz, au	XVII	e siècle, a imaginé un calcul
un sujet d'actualité au	XXI	e siècle. □
En juillet 2009, en	Californie	une conférence organisée par l
permanente dans le domaine à	Paris	. □
et de compréhension de l'	université Stanford	. □
□ En	France	□
□ En	France	, les pionniers sont Alain Colmerauer
est créé en 1979 à	Toulouse	. □
Conference on Artificial Intelligence à	Chambéry	en 1993, et la création
une audition se tient au	Sénat	: « L'intelligence Artificielle menace-t-elle
la République Emmanuel Macron au	Collège de France	pour annoncer la stratégie de
Taux de rappel	18/25	72.00 %
Taux de précision	18/30	60.00 %

Notons ici l'apparition de quelques nombres romains, qui est non désirée, ainsi que l'absence de certains noms de lieux avec une construction un peu particulière, tel que "université d'Édimbourg".

---

Deuxième itération

---

## Analyse

Parmis les erreurs que nous avons reconnues dans les résultats de la première itération, nous avons conclu qu'il était assez simplement possible d'éliminer les nombres romains grâce à une vérification sur les lettres composant les mots reconnus qui sont exclusivement en lettres majuscules, et de ne pas retenir ceux qui ne contiennent que des lettres utilisées dans l'écriture de chiffres romains.

## Algorithme et Script Python

Avec cette version nous reprenons la version 1, mais tentons d'exclure les nombres romains.

- 1) Stocker le texte dans la variable *texte*.

- 2) Couper le texte en mots distinct sur la base des espaces et de la ponctuation.
- 3) Définir les prépositions locatives introduisant typiquement des lieux (*en, au, aux, à*).
- 4) Itérer mot à mot sur le texte.
- 5) Pour chaque mot commençant par une majuscule et étant précédé d'une préposition de lieux :
  - a) Si le mot n'est pas exclusivement en majuscules, imprimer le mot.
  - b) Si le mot est exclusivement en majuscules, mais pas exclusivement composé de lettres utilisées pour dénoter un nombre romain (I, V, X, L, C, D, M), imprimer le mot.

```
import re

texte = open('IApourTP1.txt',encoding='utf8').read()
mots = re.split(r'[^\\w|-]+',texte)

prep_avant = ["en", "au", "aux", "à"]

for (idx, mot) in enumerate (mots) :
    if (len(mot) > 0 and mot[0].isupper()) :
        if (mots[idx-1].lower() in prep_avant) :
            if (not mot.isupper()) :
                print(mots[idx-5:idx+6])
            else :
                for letter in mot :
                    if not (letter in "IVXLCDM") :
                        print(mots[idx-5:idx+6])
                        break
```

## Évaluation

**Vert** → Mot cible retenu correctement

**Jaune** → Mot cible non retenu ou seulement en partie

**Rouge** → Mot retenu incorrectement

**Barre** → Mot qui était faussement reconnu dans l'itération précédente, qui ne l'est maintenant plus

marqué par une conférence au	Dartmouth College	à l'été 1956 à
recherche se fait principalement aux	États-Unis	, notamment à l'université Stanford
aux États-Unis, notamment à l'	université Stanford	sous l'impulsion de John
impulsion de John McCarthy au	MIT	sous celle de Marvin Minsky
de Marvin Minsky, à l'	université Carnegie-Mellon	sous celle de Allen Newell
Herbert Simon et à l'	université d'Édimbourg	sous celle de Donald Michie
celle de Donald Michie, en	Europe	et en Chine, ainsi qu'
Michie en Europe et en	Chine	, ainsi qu'au Japon avec
en Chine, ainsi qu'au	Japon	avec le projet « ordinateurs de
□ En	France	l'un des pionniers est
800 avant notre ère, en	Égypte	. □
de bronze Talos, gardien des	rivages de la Crète	, était parfois considéré comme une
ingénieurs fondée par Ctesibius à	Alexandrie	, et concevant des mécanismes destinés
□ On a prêté à	Roger	Bacon la conception d'automates
exposé pendant plusieurs années en	France	, en Italie et en Angleterre
plusieurs années en France, en	Italie	et en Angleterre, et la
France, en Italie et en	Angleterre	, et la transparence de l
Malheureusement, l'incendie du	musée de Nijni Novgorod	en Russie vers 1879 détruisit
musée de Nijni Novgorod en	Russie	vers 1879 détruisit cet automate.
□ Gottfried Wilhelm Leibniz, au	XVII	e siècle, a imaginé un calcul
un sujet d'actualité au	XXI	e siècle. □

En juillet 2009, en	Californie	une conférence organisée par l
permanente dans le domaine à	Paris	. □
et de compréhension de l'	université Stanford	. □
□ En	France	□
□ En	France	, les pionniers sont Alain Colmerauer
est créé en 1979 à	Toulouse	. □
Conference on Artificial Intelligence à	Chambéry	en 1993, et la création
une audition se tient au	Sénat	: « L'intelligence Artificielle menace-t-elle
la République Emmanuel Macron au	Collège de France	pour annoncer la stratégie de
Taux de rappel	18/25	72.00 %
Taux de précision	18/28	64.28 %

Les deux nombres romains présents dans le texte ont été filtrés, améliorant la précision de l’algorithme sans affecter le taux de rappel. On pourrait toutefois éventuellement imaginer qu’un nom de lieu puisse voir son sigle contenir seulement des lettres qui ont été marquées comme étant utilisées dans les nombres romains par coïncidence, mais cela est plutôt rare, et de tels sigles ne surviennent pas dans le texte ici utilisé. Également, plusieurs noms de lieux qui sont ici des noms de collèges ou d’universités n’ont pas été retenus, mais une structure telle que

[article] [type d’institution] [préposition génitive] [nom propre]  
(e.g. *l’université d’Édimbourg*)

est difficile à retenir avec un algorithme qui n’utilise pas spécifiquement des mots comme “université” ou “collège” sans sévèrement diminuer la précision. Des intrus comme “à Roger” ou “au Sénat” ne font pas référence à des lieux, mais suivent une formulation qui correspond à celle que nous ciblions. Nous croyons que ceux-ci seraient plutôt difficiles à éliminer sans une analyse sémantique beaucoup plus approfondie.



## Annexe I - Extraits

---

- l'intelligence artificielle devient un domaine de recherche international, marquée par une conférence au Dartmouth College à l'été 1956 à laquelle assistaient ceux qui vont marquer la discipline ;

- depuis les années 1960, la recherche se fait principalement aux États-Unis, notamment à l'université Stanford sous l'impulsion de John McCarthy, au MIT sous celle de Marvin Minsky, à l'université Carnegie-Mellon sous celle de Allen Newell et Herbert Simon et à l'université d'Édimbourg sous celle de Donald Michie, en Europe et en Chine, ainsi qu'au Japon avec le projet « ordinateurs de cinquième génération (en) » du gouvernement. En France, l'un des pionniers est Jacques Pitrat ;

---

En septembre 2017, Cédric Villani, premier vice-président de l'Office parlementaire d'évaluation des choix scientifiques et technologiques (OPECST), est chargé de mener une consultation publique sur l'intelligence artificielle. Il rend son rapport le 28 mars 2018, à la veille d'une intervention du président de la République Emmanuel Macron au Collège de France pour annoncer la stratégie de la France dans ce domaine. Il y dévoile un plan de 1,5 milliard d'euros sur l'ensemble du quinquennat, ainsi qu'une évolution de la législation française pour permettre la mise en application de l'intelligence artificielle, en particulier concernant la circulation des véhicules autonomes. Parallèlement à ces annonces, il est interviewé par Wired, magazine de référence pour la communauté mondiale des nouvelles technologies, et y exprime sa vision de l'intelligence artificielle, à savoir que les algorithmes utilisés par l'État doivent être ouverts, que l'intelligence artificielle doit être encadrée par des règles philosophiques et éthiques et qu'il faut s'opposer à l'usage d'armes automatiques ou de dispositifs prenant des décisions sans consulter un humain.

---