

CONCEVOIR UN SYSTÈME DE GÉNÉRATION DE TEXTE

POUR COMPRENDRE L'INTELLIGENCE ARTIFICIELLE

Le développement de systèmes de génération de texte est un sujet d'actualité de la recherche en informatique. C'était l'objectif qu'avait fixé Turing dans son jeu de l'imitation pour qu'une intelligence artificielle puisse être dite véritablement intelligente. Pour en comprendre les enjeux et les limites, rien de tel que de concevoir concrètement un tel système.

Objectif

La conception concrète du système de génération de texte permet de bien sentir les mécanismes utilisés actuellement, même dans des systèmes très avancés. Via cet exemple simple, on comprend leur fonctionnement, tout en constatant les limites. Cela permet d'introduire les systèmes constituant l'état de l'art.

Fiche technique

PUBLIC : 12 ans et plus.

ACTIVITÉS : construction et exploration d'un graphe.

MATÉRIEL NÉCESSAIRE : un vidéo-projecteur avec un ordinateur, éventuellement un paperboard ; un exemplaire de *Huis-Clos*, de Jean-Paul Sartre.

TEMPS CONSEILLÉ : 1h.

Une fiche proposée par
L'Atelier

Médiathèque des Jardins de la
Culture

Contact : Yohan Chalier
y.chalier@rlv.eu

DÉROULÉ

15 min : Introduction

- Présentation de l'objectif de la séance.
- Questionnement mené par l'animateur : « Comment créer un tel système de façon à ce qu'il soit exploitable par un ordinateur ? ». L'animateur guide les participants vers la notion d'algorithme.
- Introduction des systèmes de langages, et de leur nature probabiliste, via l'exemple de la prédiction du clavier de téléphone.

15 min : Conception du modèle

- Présentation de la représentation d'un modèle de langage en graphe.
- En utilisant le texte de la Scène 2 de l'Acte I de *Huis-Clos*, création d'un proto-modèle en dessinant sur un paperboard ou sur un écran projeté. Le modèle se base sur l'extraction d'une relation simple : on compte le nombre de fois que le mot *X* suit le mot *Y*.
- Lorsque cette tâche devient trop complexe pour être écrite clairement par un humain, on montre le résultat de l'entraînement sur une machine depuis l'interface en ligne.

10 min : Généralisation

- Présentation de l'interface en ligne : nœud central, nœuds extérieurs, épaisseur des arêtes, cas particulier du point.
- Génération de phrases.

20 min : Ouverture

- Mise en évidence des artefacts de l'entraînement (par exemple, les phrases commencent souvent par « *garcin* », « *estelle* », « *inès* ») et discours général sur les biais des algorithmes.
- Mise en évidence des limitations (beaucoup de phrase absurdes) et introduction à la nécessité d'extraire des relations plus profondes.

RESSOURCES

- Article de vulgarisation sur les modèles de langage :
<https://atelier-mediathèque.rlv.eu/blog/une-histoire-dauteur>
 - Interface en ligne :
<https://dir.chalier.fr/public/huis-clos/>
-