TD 3 - Information Extraction and Knowledge Graph Population

Objectif principal

L'objectif de ce TD est double : d'une part, comprendre en profondeur le fonctionnement des algorithmes et modèles d'extraction d'informations (IE) et, d'autre part, apprendre à les utiliser pour créer et enrichir un graphe de connaissances.

Outils et modèles

- **Modèle REBEL** : Nous utiliserons le modèle REBEL (disponible sur GitHub) pour l'extraction d'entités et de leurs relations à partir de textes. REBEL est un modèle de pointe pour l'extraction d'informations relationnelles.
- **Bibliothèque Python "Transformers"** : Pour interroger le modèle REBEL, nous utiliserons la bibliothèque "transformers". Cette bibliothèque facilite le chargement et l'utilisation de modèles de traitement de langues naturelles comme REBEL, directement depuis Huggingface.
- **Bibliothèque Python NetworkX** : Les résultats de l'extraction seront organisés et sauvegardés sous forme de triples en utilisant NetworkX, une bibliothèque pour la création, la manipulation et l'étude de la structure, de la dynamique et des fonctions des réseaux complexes. Vous pouvez trouver plus d'informations et des tutoriels sur NetworkX ici.

Procédure

- **Découpage** : Le texte source sera traité en plusieurs blocs pour faciliter l'analyse et l'extraction d'informations.
- **Extraction d'entités et de relations** : Utilisez REBEL pour identifier les entités clés et leurs relations interconnectées dans chaque bloc de texte.
- **Création du graphe de connaissances** : Organisez les entités et relations extraites sous forme de graphe à l'aide de NetworkX. Cela implique la création de nœuds (entités) et d'arêtes (relations) pour visualiser les liens entre les éléments du texte.
- **Analyse et Interprétation** : Après la création du graphe, analysez-le pour comprendre comment les informations sont interconnectées et ce qu'elles révèlent sur le contenu du texte.

Pour aller plus loin

Trouver quel type d'algorithme met en œuvre REBEL et comment son dataset est structuré.

Geoffrey Everest Hinton (born 6 December 1947) is an English Canadian cognitive psychologist and computer scientist, most noted for his work on artificial neural networks. Since 2013 he divides his time working for Google (Google Brain) and the University of Toronto. In 2017, he cofounded and became the Chief Scientific Advisor of the Vector Institute in Toronto.

With David E. Rumelhart and Ronald J. Williams, Hinton was co-author of a highly cited paper published in 1986 that popularized the backpropagation algorithm for training multi-layer neural networks, although they were not the first to propose the approach. Hinton is viewed by some as a leading figure in the deep learning community and is referred to by some as the "Godfather of Deep Learning". The dramatic image-recognition milestone of the AlexNet designed by his student Alex Krizhevsky for the ImageNet challenge 2012 helped to revolutionize the field of computer vision. Hinton was awarded the 2018 Turing Prize alongside Yoshua Bengio and Yann LeCun for their work on deep learning.

Hinton was educated at King's College, Cambridge graduating in 1970, with a Bachelor of Arts in experimental psychology. He continued his study at the University of Edinburgh where he was awarded a PhD in artificial intelligence in 1978 for research supervised by Christopher Longuet-Higgins.

After his PhD he worked at the University of Sussex, and (after difficulty finding funding in Britain)[26] the University of California, San Diego, and Carnegie Mellon University. He was the founding director of the Gatsby Charitable Foundation Computational Neuroscience Unit at University College London, and is currently a professor in the computer science department at the University of Toronto. He holds a Canada Research Chair in Machine Learning, and is currently an advisor for the Learning in Machines & Brains program at the Canadian Institute for Advanced Research. Hinton taught a free online course on Neural Networks on the education platform Coursera in 2012. Hinton joined Google in March 2013 when his company, DNNresearch Inc., was acquired. He is planning to "divide his time between his university research and his work at Google".

Hinton's research investigates ways of using neural networks for machine learning, memory, perception and symbol processing. He has authored or co-authored over 200 peer reviewed publications.

While Hinton was a professor at Carnegie Mellon University (1982–1987), David E. Rumelhart and Hinton and Ronald J. Williams applied the backpropagation algorithm to multi-layer neural networks. Their experiments showed that such networks can learn useful internal representations of data. In an interview of 2018, Hinton said that "David E. Rumelhart came up with the basic idea of backpropagation, so it's his invention." Although this work was important in popularizing backpropagation, it was not the first to suggest the approach. Reverse-mode automatic differentiation, of which backpropagation is a special case, was proposed by Seppo Linnainmaa in 1970, and Paul Werbos proposed to use it to train neural networks in 1974.

During the same period, Hinton co-invented Boltzmann machines with David Ackley and Terry Sejnowski.[32] His other contributions to neural network research include distributed representations, time delay neural network, mixtures of experts, Helmholtz machines and Product of Experts. In 2007 Hinton coauthored an unsupervised learning paper titled Unsupervised learning of image transformations. An accessible introduction to Geoffrey

Hinton's research can be found in his articles in Scientific American in September 1992 and October 1993.

In October and November 2017 respectively, Hinton published two open access research papers on the theme of capsule neural networks, which according to Hinton are "finally something that works well".

Notable former PhD students and postdoctoral researchers from his group include Richard Zemel, Brendan Frey, Radford M. Neal, Ruslan Salakhutdinov, Ilya Sutskever, Yann LeCun and Zoubin Ghahramani.