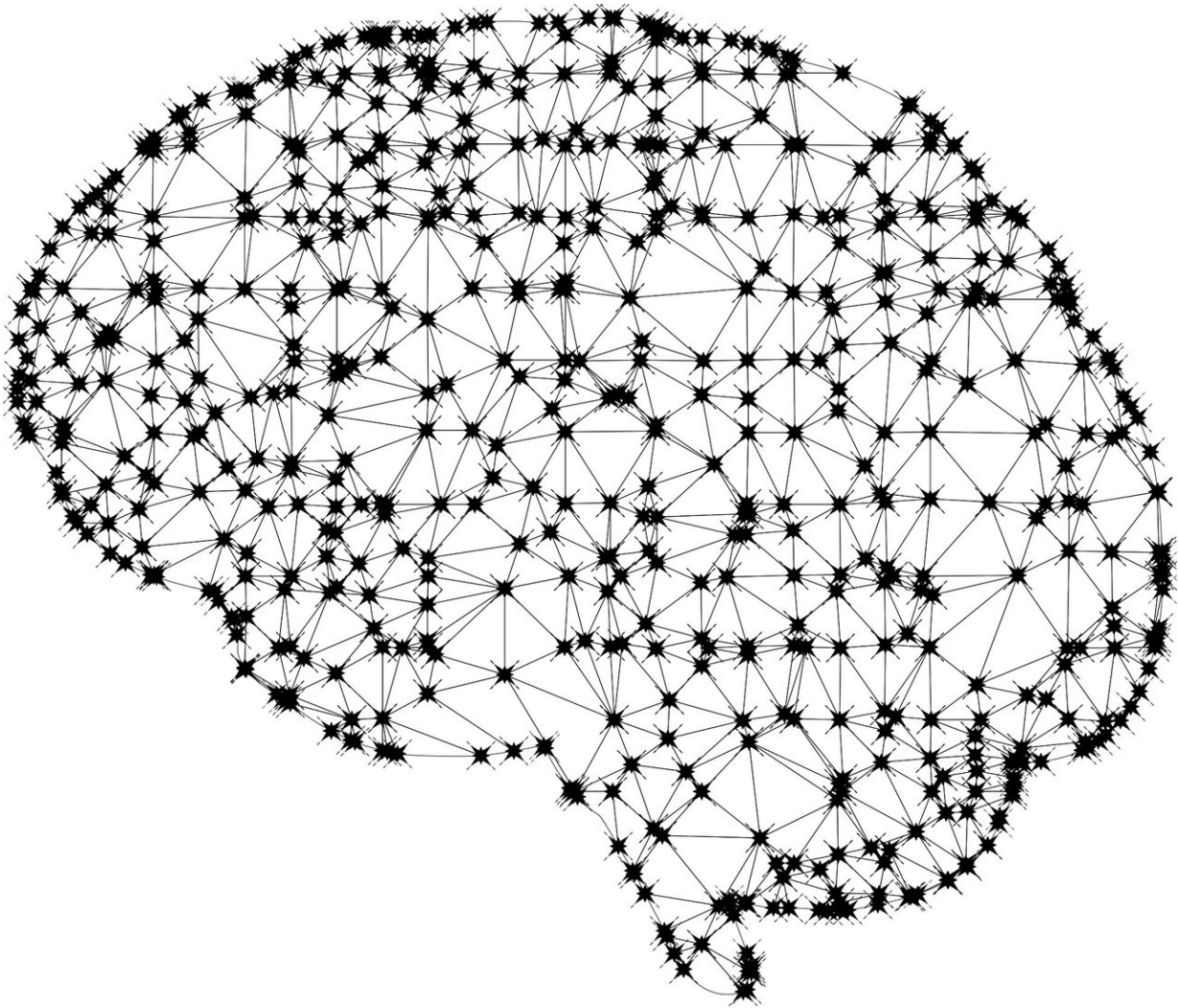# New technology for machine translation now available

January 22 2019

Credit: CC0 Public Domain

A new methodology to improve machine translation has become available this month through the University of Amsterdam. The project DatAptor, funded by NWO/STW, increasingly advances translation machines by selecting data sets.

The methodology is used in the application Matching Data, offered by TAUS, an important think tank in the field of [machine translation](). This application tackles a big challenge within digital [translation](): for a good translation it is necessary to train the translation machine with reliable sources and datasets that contain the relevant type of words. For example, translating a legal text requires a completely different vocabulary and a different type of translation than for example, a newspaper report.

Successful implementation

In 2013 the DatAptor project, supervised by Professor Khalil Sima'an of the UvA Institute for Logic, Language and Computation, received funding from Technology foundation STW (now: NWO Domain Applied and Engineering Sciences) to deal with this problem. The research results of the DatAptor project have now been successfully implemented by think tank TAUS. They offer the new technology under the name Matching Data.

On the weblog of TAUS Sima'an says: "Our dream was to make the world wide web itself the source of all data selections. But we decided to start more modest and make the very large TAUS Data repository our hunting field first. In DatAptor we learned that every domain is a mixture of many subdomains. The combinatorics of subdomains in a very large repository harbors a wealth of new, untapped selections. Therefore, if the user provides a Query corpus representing their domain of interest, the Matching Data method is likely to find a suitable selection in the repository."

More information: Data-Powered Domain-Specific Translation Services On Demand (DatAptor)