The Google BERT hype - is it really the biggest leap forward in the past five years?

Recent weeks have witnessed an immense hype about the new Google BERT algorithm. This is presented as a huge leap forward in text understanding by machines and the new paradigm for semantic technologies. While any AI advance and new search update by Google is always welcome, we would like to point out that the supposed new capabilities of BERT are NEITHER UNIQUE NOR NEW! These supposed "new capabilities" of the BERT algorithm have been obtained and proved in an entirely different approach by the InfoCodex Semantic Engine years ago. This algorithm was indeed the first to demonstrate its power for drug discovery based only on AI text analysis on very large document corpora in a concrete test in collaboration with Merck and Thomson Reuters, something that has always eluded other, more hyped products for AI in health care.

BERT is an NLP product designed mainly for English. By its very definition, NLP understands relations in human language and is thus focused only on information retrieval and search. The "holy grail" of semantic AI, however is knowledge discovery rather than search and retrieval. The really difficult task is to find NEW information that is NOT written explicitly in human text by finding hidden correlations. This is what InfoCodex can do.

Disambiguation tasks, like the much hyped recognition of the different meanings of the word "bank" in the expressions "river bank" and "bank account", have been solved many years ago by other methods than BERT, nothing new here. Moreover, while BERT is pre-trained on a large corpus, it still has to be trained on the specific tasks it should be applied to. So, even the training problem has not been solved, something that InfoCodex has, instead, solved years ago. Finally, it is claimed that BERT can be generalized to other languages. But again, many separated languages is not the real problem, the real difficult problem is to obtain retrieval and discovery ACROSS languages, something that InfoCodex has also solved.

BERT is a great update to the Google search algorithm. It is not, however, the hyped breakthrough in semantic AI that the press is swarming about. Most of the claimed "advances" are present in commercial products like InfoCodex since years.