

NOOSPEAK

by Matteo Menapace (iosonomatteo@gmail.com)

[NooSpeak](#)¹ aims to enhance Web search with human-like interaction, in order to provide users with a more natural and entertaining experience. It merges an existing widely used service ([Google Search](#)) with AI² technology (a pattern knowledge-base written in [AIML](#)³ and its parser) to create a *chatting* interface to the search engine.

The project starts from the assumption that search is highly computer-centred, forcing people to formulate their questions in an unnatural form, and produces a poor and frustrating experience. The goal of the project is therefore to build an interaction system that facilitates users by adopting their natural language, yet preserves the efficiency of the search engine in finding information on the Web. The first prototype of such a system uses a chatterbot⁴ to pre-process human inputs, passing the chatterbot responses to the search engine, from which results it extracts one (the most relevant⁵) to be delivered to the user. Search results are then presented inside an interface that borrows from instant messaging software conventions, so that user has a constant overview of her chatting/searching history. The chatterbot role is a diagnostic one, helping users to clarify and refine their search, possibly correcting spelling mistakes.

In order to ease user's task, the prototype offers real-time suggestions⁶ based on popular searches relevant to what user is typing.

It also displays two weighted lists⁷ (*chat clouds*) of semantically related⁸ words and synonyms, one gathering data from all users inputs and the other built upon the corpus of Google responses. Chat clouds' function is to enrich user's vocabulary, aiding her to find new solutions and build relationships between topics.

A further stage of the project aims to allow users to customize the service: by creating an account, they would be able to actively teach the chatterbot about their preferences, and to see the system improving its performances by analysing their individual data.

WHY SEARCH SUCKS

The enormous success of Google reveals how search is becoming the interface that an increasing number of people use to navigate the Web's rapidly expanding amount of information. It has become an universally understood method, and the best solution so far to the problem of avoiding the equation "infinite = null", by which the information someone is looking for exists in the Web but is not available, because she is not able to find it. regarding information on the Web. Yet, while huge efforts and budgets have been devoted

¹ the name NooSpeak is a reference to Newspeak, the fictional language used by Party members in the dystopian novel "1984" by George Orwell. Ideally, it evokes a new language for computers that simplifies interaction with humans.

² Artificial Intelligence

³ Artificial Intelligence Markup Language, an XML derivate developed by Richard Wallace, creator of the open-source [A.L.I.C.E.](#) project.

⁴ a software designed to simulate an intelligent conversation, which scans for keywords within the input and pulls a reply with the most matching keywords or the most similar wording pattern from a local database.

⁵ the definition of relevance is arguable, but for the purpose of this prototype it refers to a sentence within the search engine results that sticks at best to the chatterbot response, and has the highest [PageRank](#)

⁶ functionality not yet implemented, see [Google Suggest](#) for an example of the same process applied to a keyword-based language

⁷ commonly known as tag clouds, they are a visual representation of websites' tags, usually sorted by popularity, alphabet and/or actuality.

⁸ functionality not yet implemented in the prototype

to improve the data mining and indexing part of a search process (the back end of a search engine), not enough attention has been given to user experience. Users are averagely considered *lazy*. According to a 2004 [Majestic](#) report⁹, nearly 50 percent of all searches use two to three words, and 20 percent use just one; just 5 percent of all searches use more than six words. A [Pew research](#)¹⁰ shows that the average number of searches per visit to an engine is nearly five: this implies that users are not getting what they want the first time, or they are formulating new requests driven by the results their initial questions return. Perhaps it is not just about laziness (or, to put it in its positive form: a search for simplicity), and a lack of comfort with a technology-driven language and its consequently frustrating user experience could better explain these facts. Search can be divided into two types, depending on its goals: a search for recovery (finding that which one knows exists) and a search for discovery (finding what one intuitively exists, or desires). In the first type user has clearly in mind what she is looking for, and the biggest barrier between her and the wished result is the right combination of keywords. In the second case user is more open to a wide array of results¹¹, from which she will refine or expand her search. In both cases, the search process can benefit from a natural language approach.

WHY AI SUCKS

A.L.I.C.E.¹², from which NooSpeak's chatterbot is derived, is a form of so called weak AI¹³: its software uses a customisable database of patterns to form sentences, but it has no reasoning and cognitive abilities. In other words, it is an efficient problem-solver, but it does not assign any semantic value to user inputs. Its limited database works well for specific fields, such as customer service Web services¹⁴, requiring specific and limited information: clearly, this is not the case of the Web. As John Battelle states, "*search will more likely become intelligent via the clever application of algorithms that harness and leverage the intelligence already extant on the Web – the millions and millions of daily transactions, utterances, behaviours, and links that form the Web's foundation – the Database of Intentions*"¹⁵.

There is no need for strong AI, but rather for the design of systems that allow user to actively (and in some cases, passively) produce semantic relationships between data that AI cannot make sense of.

CHAT+SEARCH=NOOSPEAK

Internet is a *dumb network*, and it works because it simply transports packets of data without needing to know anything about them, and the end devices (computers, for example) contain the intelligence.

Similarly, Google is efficient and trustable because it does not judge/edit the Web, but relies on the collective intelligence of millions of users, and their judgements (links) to organise it.

⁹ as cited by Battelle 2005, *The Search*, pg. 27

¹⁰ The Pew Internet & American Life Project produces reports that explore the impact of the Internet on families, communities, work and home, daily life, education, health care, and civic and political life in the USA.

¹¹ although 50 percent of users do not go beyond the first ten results, as stated by Spink and Jansen 2004, [A study of Web Search Trends](#)

¹² [A.L.I.C.E.](#) is an open-source, artificial intelligence natural language chat robot.

¹³ for more information on the topic, see Wikipedia's [Strong AI vs. Weak AI](#)

¹⁴ an example of this approach is [Ikea's Anna chatterbot](#)

¹⁵ Battelle 2005, *The Search*, pg. 17

The aim of NooSpeak is not to create an intelligent entity, but a system that effectively connects people in a more human way (user-centred design). It uses human language, but it does not try to emulate human appearance.