Software Project

Answer real questions using Wikipedia content

Frederic Assmus, Axel Didier, Pierre Gonçalves

GOAL

Identify the answers to real user questions about Wikipedia page content

Use, understand and improve a question-answering model based on Wikipedia articles. Given a real human question about a Wikipedia article, we should predict 2 answers : one short and one long. These answers have to be relevant and retrieved from the article.

https://www.kaggle.com/c/tensorflow2-question-answering/overview

Brewing

From Wikipedia, the free encyclopedia

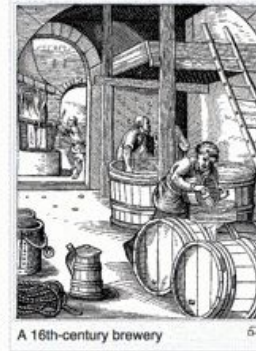> "Brewer" redirects here. For other uses, see Brewer (disambiguation).
> This article is about the brewing of beer. For homebrewing, see Homebrewing. For other uses, see Brewing (disambiguation).

Brewing is the production of beer by steeping a starch source (commonly cereal grains, the most popular of which is barley)[1] in water and fermenting the resulting sweet liquid with yeast. It may be done in a brewery by a commercial brewer, at home by a homebrewer, or by a variety of traditional methods such as communally by the indigenous peoples in Brazil when making cauim.[2] Brewing has taken place since around the 6th millennium BC, and archaeological evidence suggests that emerging civilizations including ancient Egypt[3] and Mesopotamia brewed beer.[4] Since the nineteenth century the brewing industry has been part of most western economies.

The basic ingredients of beer are water and a fermentable starch source such as malted barley. Most beer is fermented with a brewer's yeast and flavoured with hops.[5] Less widely used starch sources include millet, sorghum and cassava.[6] Secondary sources (adjuncts), such as maize (corn), rice, or sugar, may also be used, sometimes to reduce cost, or to add a feature, such as adding wheat to aid in retaining the foamy head of the beer.[7] The proportion of each starch source in a beer recipe is collectively called the grain bill.

Steps in the brewing process include malting, milling, mashing, lautering, boiling, fermenting, conditioning, filtering, and packaging. There are three main fermentation methods, warm, cool and spontaneous. Fermentation may take place in an open or closed fermenting vessel; a secondary fermentation may also occur in the cask or bottle. There are several additional brewing methods, such as barrel aging, double dropping, and Yorkshire Square.

A 16th-century brewery

Contents [hide]

Question : When are hops added to the brewing process?

Difficulties to come

- **Answers format** : For a given question, the answer can have 3 different formats : a set of token indices (corresponding to the range of the tokens in the wikipedia article) OR Yes or No (only for short answers, of course) OR a blank if we couldn't find any answer. So we see that short answers can be YES/NO answers when it is possible, or can be sentences when YES/NO is not applicable (= open questions). So **one of our task would be to find if questions are open or not.**

- Going from **oriented to naive** question : In the dataset, questions and Wikipedia articles are already linked together. A future interesting problem would be, for a given question, to determine in **which Wikipedia article** the answers have to be found.

- **Performance ?** We'll have to evaluate our performance using the Kaggle statement.

Tools

- 16,27 Gb of textual Data, split in train / test (JSON Lines)

- Data Fields : Article text, question, array containing the short answer + the long answer, ….

- Word embeddings : Word2Vec / **BERT embeddings**

- Baseline : existing notebooks from Kaggle competitors

References

- https://www.kaggle.com/c/tensorflow2-question-answering/overview/description

- https://ai.google.com/research/NaturalQuestions/visualization

- https://www.quora.com/What-are-the-main-differences-between-the-word-embeddings-of-ELMo-BERT-Word2vec-and-GloVe

- https://medium.com/@dhartidhami/understanding-bert-word-embeddings-7dc4d2ea54ca