

Using NLP to deliver relevant content in Online Publishing

Getting paid for content in the days of blogs and social networks is tough! There is a lot of competition and just producing high-quality content is no longer enough. In order to attract advertisers and get digital subscriptions, online publishers need to make sure that they have an engaged audience. To create an engaged audience, it is crucial that relevant content reaches the right people. This white paper post highlights the importance of Natural Language Processing in this cycle.

Recent research shows that [publishers are increasingly relying on data about their readers to drive engagement](#), reduce churn and enhance ad sales. So, in order to provide the most interesting and relevant content to readers, it is important to first of all collect data about them. What content do they read, like and comment on? What can we find out about them? Perhaps

they have shared their Facebook or LinkedIn page, which in turn contains information about their interests and circles of friends.

Once we have collected the data, the interesting but also challenging part starts. Interesting, because this data contains answers to questions like “What type of content is most likely to keep this type of reader on my site?”. Challenging, because the majority of this data is in textual form, unstructured and difficult to generalize.

Natural Language Processing can help with these challenges. One way of generalizing content is to apply topic indexing, which reduces text down to a consistent set of key topics. These topics are then used to automatically determine similar content. Each reader is interested in certain topics, which they don't usually disclose, but these topics can be inferred through the analysis of the content they read or the information they reveal about themselves. Deriving such topics makes it easy for us to understand what's happening in the data, but sometimes it is difficult to capture the essence of a particular group of articles or users in terms of topics. Here, a different NLP technique comes in place, called classification or clustering. It can capture the underlying similarities between different content and between readers' profiles without having to name them, and it can tell us things like: this user belongs into this group of our readers; people in this group have enjoyed this kind of content.

Independently from the technique, the key point is that NLP and text analytics can help online publishers build an engaged audience by recommending interesting content to their readers. Some publishers may benefit from a general content recommender, like the one [Google is planning to release soon](#). Others, may need a custom solution. Entopix has been recently engaged for exactly this kind of application, although unfortunately at this stage

the customer's name may not be revealed due to the competitiveness of this market.