



ALLAN THRAEN |

🕒 15 years ago |

📄 PDF |

💬

AUTO-TAGGING CONTENT WITH OPENCALAIS



One of my big passions has always been various forms of intelligent textual analysis – probably a remnant of my search-engine days. Anyway – Over the last couple of years I've done a bunch of different prototypes which have it in common that they focus on working with the content part of content management. That is, they look at the actual text editors insert – and try to do something semi-clever with it. I've done prototypes that expand the links collection property-type with a “Suggest Related” button, stuff that attempts to suggest meta-tag keywords automatically, etc. Most of this I never got around to polish off and blog – it's mainly been fun little experiments. One of those was done 8-9 months ago, when I was playing around with OpenCalais one day. OpenCalais is an awesome service that allows you to submit textual content and then it tags it – and extract key parts of the information – like which people, companies, technologies, etc that are mentioned in the text. I know I haven't been alone in playing around with OpenCalais – but I figured I'd share my basic prototype here.

My initial idea was to make a “hidden” connection to OpenCalais – so whenever a page was published, it would automatically be tagged – and the tags would be stored in the categories. However this turned out to be quite a slow procedure so I abandoned that idea. My current approach is to have a tab in edit-mode where you can ask to have the page analyzed – and if you like the result, you can push a button and save that into the categories.

If you choose to save it to categories, it will be setup in a category structure.

The code is still “very prototypish” so unless I hear a great, instant demand I'm not planning to release the code now. If anybody want to play around with the binaries, you are welcome – just unzip this package and copy to your site folder. Works in both CMS 5 and CMS 6. NOTE: This is a prototype, not production ready code. But feel free to be inspired – this was done in an afternoon – imagine what you can do with it in a full day.

CodeArt ApS

Teknikerbyen 5, 2830 Virum, Denmark

Email: info@codeart.dk

Phone: +45 26 13 66 96

CVR: 39680688

