

# Beyond Google

by Nils Mulvad

## Training session at the global investigative journalism conference in Geneva 23th of April 2010

*By nils.mulvad@kaasogmulvad.dk*

Google only search a very limited part of the web. Beyond Google you have metadata, deleted websites and all content in searchable databases. And then you also have a lot of other search-engines focused on specific tasks.

Metadata: hidden information in files. It can be date of construction, but much more.

Deleted websites: Go to Archive.org to find them. Type in the domain or the direct link to a website and find older versions.

Searchable databases: The list is enormous. Here are some examples:

Medline – use this for scientific information: <http://www.ncbi.nlm.nih.gov/PubMed/>

[www.tobaccoarchives.com](http://www.tobaccoarchives.com) – find old internal documents from tobacco companies. Many of these databases uses boolean logic as search language.

Examples of search strings using boolean logic in the Philip Morris database:

Pauthor:voss,t and (Ptype:(invoice OR budget OR Pay Request) OR Otype:(invoice OR budget OR Pay Request))

(Pmentioned:(carlson,s OR martin,p OR reif,h) OR recipient:(carlson,s OR martin,p OR reif,h)) and (Ptype:(invoice OR budget OR Pay Request) OR Otype:(invoice OR budget OR Pay Request))

(Pmentioned:(carlson,s OR martin,p OR reif,h) OR recipient:(carlson,s OR martin,p OR reif,h)) AND characteristics: confidential

Dutch recipients of farmsubsidies:

[http://www.minlnv.nl/portal/page?\\_pageid=116,3595803&\\_dad=portal&\\_schema=PORTAL](http://www.minlnv.nl/portal/page?_pageid=116,3595803&_dad=portal&_schema=PORTAL)

Searching into this shows us the authorities making a lot of spelling errors and trying to make it difficult to get an overview of the total payments to each recipient. Here a NGO, farmsubsidy.org, extract and structure the data, and they will be happy to share the data with journalists. You see a growing environment for this.

Another search engine:

<http://search.twitter.com/advanced>

See a [handout](#) on the use of Twitter