



# Are you master of your thesis? ARC seminar 2014-02-06

Göran Hamrin, PhD

Logician Lecturer Librarian

(with an idea from Bengt Haraldsson)



WHAT is this?





The manufacturer (The "author") has decided to call it a

**CHOPPOMOWER**



Suppose that the object in the picture above represents a scientific article.



Suppose that the object in the picture above represents a scientific article.

How would you do to search for it?



What we need is:



What we need is:

- A general method that works



## What we need is:

- A general method that works
- And where to find additional help





But before we search we need to know:



But before we search we need to know:

How is scientific information collected,  
indexed, saved and made searchable?



Bibliografic reference (to a scientific paper/etc)



## Bibliographic reference (to a scientific paper/etc)

Is divided into different searchable **fields**



## Bibliografic reference (to a scientific paper/etc)

Is divided into different searchable **fields**

Title field: Here you find the title of the paper (big surprise 😊)

Author field: names of all the authors

Date field: Year (date) of publication

Source field: Here you find information of where/by which organisation the paper/scientific result has been communicated/published)



## Bibliographic reference (to a scientific paper/etc)

Is divided into different searchable **fields**

Title field: Here you find the title of the paper (big surprise 😊)

Author field: names of all the authors

Date field: Year (date) of publication

Source field: Here you find information of where/by which organisation the paper/scientific result has been communicated/published)

Abstract field: contains a summary of the results/findings

...and there are more fields....



## Bibliographic reference (to a scientific paper/etc)

Is divided into different searchable **fields**

Title field: Here you find the title of the paper (big surprise 😊)

Author field: names of all the authors

Date field: Year (date) of publication

Source field: Here you find information of where/by which organisation the paper/scientific result has been communicated/published)

Abstract field: contains a summary of the results/findings

...and there are more fields....

A bibliographic reference can therefore be seen as a **vector**, and a collection of such vectors is called a **reference database**



What would a bibliographic reference to the Choppomower "article" look like?





Title field: I've just built a choppomower dude!

Author field: Chip Chipson

Date field: 2014

Source field: The pimp-my-ride magazine, vol.  
1:2014

Abstract field: I started out building the choppomower three years ago and I tell you; it's been the best project ever. It has 2000 horse powers and can go from zero to "call my mommie" in under ten seconds. It cuts the grass to a paste and all the ladies love it.

What do you remember from  
the vector (sorry, the  
"bibliographic reference")?

What do you remember from the vector (sorry, the "bibliographic reference")?

But what has Chip actually invented?  
What interesting information is included in the reference?



Well, Chip has actually  
invented a new sort of rotor blade  
which could be useful in my research  
project, since I am looking for a method  
that, swiftly and smoothly, can  
cut grass to a paste/pulp.



Well, Chip has actually  
invented a new sort  
which could be used in a  
project, since it's a very research  
that... for a method  
smoothly, can  
to a paste/pulp.

*But how should we be able to recognize  
that, looking only at a glance at the text?*



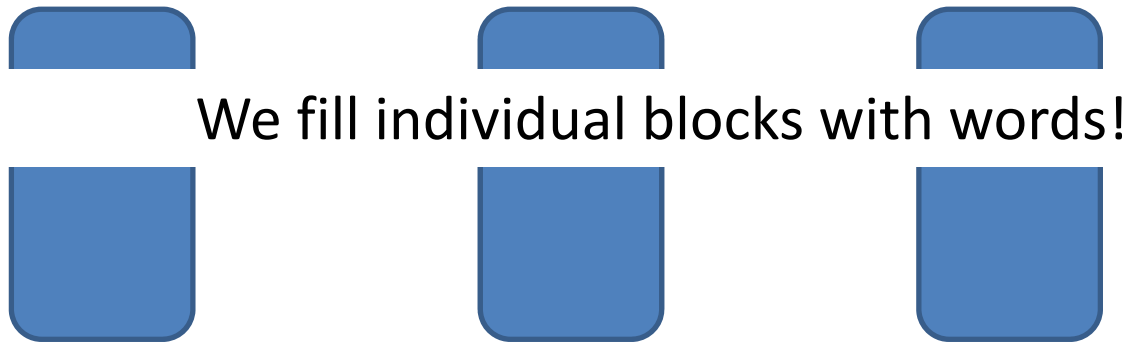
Chip has active  
invertible sort  
database of typically 10 000 000 vectors  
which could be used by research  
project, since it is for a method  
that can  
But how should we find this paper in a reference  
that, looking only at the text??

Our desire: We need a general method suitable for searching in large reference databases, given that we need information on a scientific topic/open-ended question (for example: topic of our master thesis...)



This method usually works in a reference database:

## The Building blocks strategy:







This method usually works in a reference database:

The Building blocks strategy:



But what words should we use?





First we build individual blocks:

Lawn  
Grass  
Green  
Weed  
Weeds

Cutting  
Cuts  
Mow  
Mower  
Mowing

Paste  
Sludge  
Fine  
Pulp



Then combine the words in the individual blocks:

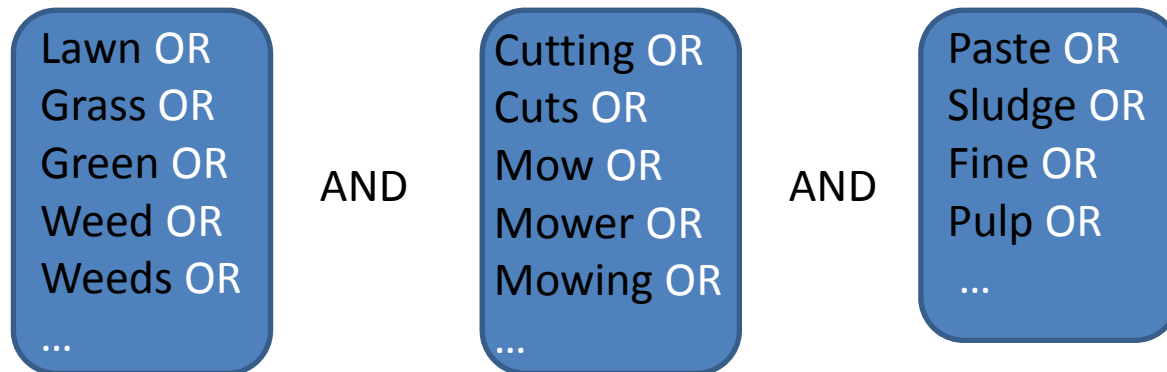
Lawn OR  
Grass OR  
Green OR  
Weed OR  
Weeds OR

Cutting OR  
Cuts OR  
Mow OR  
Mower OR  
Mowing OR

Paste OR  
Sludge OR  
Fine OR  
Pulp OR



And last, we combine the different blocks! (And repeat the process...)





Title field: I've just built a choppomower dude!

Author field: Chip Chipson

Date field: 2014

Source field: The pimp-my-ride magazine, vol. 1:2014

Abstract field: I started out building the choppomower three years ago and I tell you; it's been the best project ever. It has 2000 horse powers and can go from zero to "call my mommie" in under ten seconds. It **cuts** the **grass** to a **paste** and all the ladies love it.



To summarize: we benefit  
from systematically  
searching in, and reading,  
the abstract field!



## But what about this alternative?

Let someone else, an expert in our topic read the paper/article!

Then the expert tells us what the article is about by adding keywords that describe the paper in a new search field (so all vectors in the reference database are given an additional position)



## This alternative requires:

A controlled WORD LIST, with well-defined /standardised words which are hierarchically ordered/related to each other.

Note that these words are controlled, so the expert may not use any other words!

This is called the database's THESAURUS





## Bibliografic reference

Is divided into different searchable **fields**

Title field: Here you find the title of the paper (big surprise 😊)

Author field: names of all the authors

Date field: Year (date) of publication

Source field: Here you find information of where/by which organisation the paper/scientific result has been communicated/published)

Abstract field: contains a summary of the results/findings

**Thesaurus field:** contains the expert's keywords describing the paper



Title field: I've just built a choppomower dude!

Author field: Chip Chipson

Date field: 2014

Source field: The pimp-my-ride magazine, vol. 1:2014

Abstract field: I started out building the choppomower three years ago and I tell you; it's been the best project ever. It has 2000 horse powers and can go from zero to "call my mommie" in under ten seconds. It **cuts** the **grass** to a **paste** and all the ladies love it.

Thesaurus field: rotor blade, lawn mower, paste maker, original design



## Checkpoint: Why is it necessary to have standardised terms (in the Thesaurus field)?

1. Because the experts (the "indexers") have so many articles to read and it would take them too much time to find good words by themselves.
2. Because otherwise the same problem repeats itself; we cannot predict which words that have been used to describe an article.
3. Because otherwise an indexer would like to use all words in an article (equivalent to full text search).

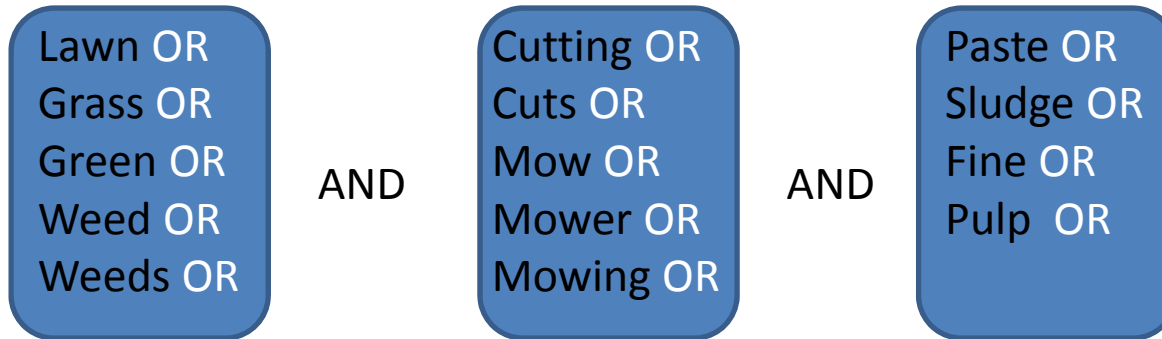


## Checkpoint: Why is it necessary to have standardised terms (in the Thesaurus field)?

1. Because the experts (the "indexers") have so many articles to read and it would take them too much time to find good words by themselves.
2. Because otherwise the same problem repeats itself; we cannot predict which words that have been used to describe an article.
3. Because otherwise an indexer would like to use all words in an article (equivalent to full text search).



We can (free keyword-/BBS-) search:

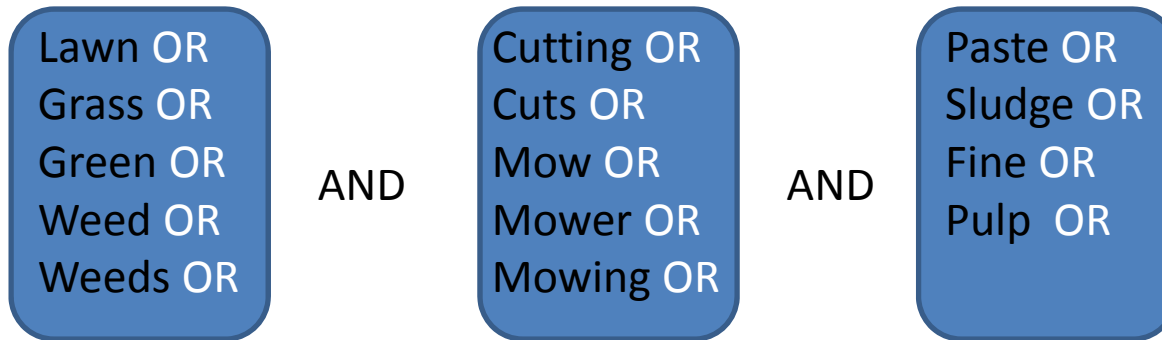


or we can (thesaurus) search:





We can (free keyword-/BBS-) search:



or we can (thesaurus) search:





That was **how** we can  
search for information.

We now talk a little of **where**  
we can search for information!



Why search?





# Why search?

- Academics
- Market value
- Ego



# Why search?

- Academics
- Market value
- Ego



- Culture



# Why search?

- Academics
- Market value
- Ego



- Culture
- Argumentation

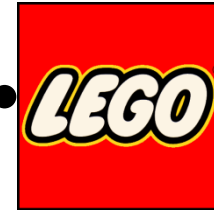


# Why search?

- Academics
- Market value
- Ego



- Culture
- Argumentation



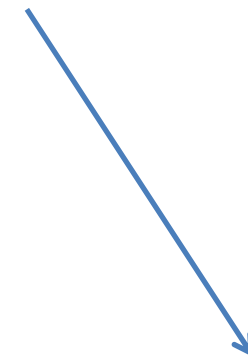
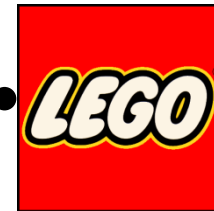


# Why search?

- Academics
- Market value
- Ego



- Culture
- Argumentation



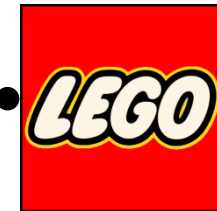
- Who is first?
- Who gets cred?
- Who gets money?



# Why search?

- Academics
- Market value
- Ego

- Culture
- Argumentation



- Read **different sources** and learn more, get a more balanced view!
- Read **more** and state your arguments more strongly (in your thesis)!
- Read!** If you do not have the knowledge - **emulate it** by reading!

- Who is first?
- Who gets cred?
- Who gets money?



# Where to search?

The **KTH Primo** database  
(is a "Google Scholar for KTH full text" and)  
can be used for:

- Finding one paper (that you know exists)
- Quick and dirty-searching
- Finding the subject-specific databases
- Finding (e-)books

Let me show you: [www.kth.se/kthb](http://www.kth.se/kthb)



# Alternatives:

## Subject-specific databases at the **KTH Library**:

- 100-200(?) databases
- Ranging från small full-text archives to large high-quality manually indexed reference databases
- "There is more than one for everyone."





# 10 KTH Schools – 10 database suggestions

”Smörgåsbord” – you try!

- One for each school
- (rather than one for each subject)



# ABE – Compendex

Civil engineering,  
Process engineering,  
Mechanical engineering,  
And much more...



# BIO - Reaxys

Substances,  
Reactions,  
Syntheses,  
Literature



# CHE – SciFinder

The comprehensive resource from  
American Chemical Society  
Natural language queries!



# CSC – ACM Digital Library

Association of  
Computing  
Machinery  
Full-text archive



ECE - ERIC

Education-related material  
U.S. Department of Education



# EES - Inspec

Physics,  
Electrical Engineering,  
And much more



# ICT - Inspec

Physics,  
Control theory,  
Information technology,  
And much more...





# ITM – Business Source Elite

Business,  
Management,  
Organization,  
Economics,  
Company profiles



# STH – PubMed/Medline

Medicine, medical technology  
and more from  
US National Library of Medicine



# SCI - MathSciNet

Mathematics, statistics, theoretical  
computer science and more



## More help?

- Take one of our courses in information searching:
- LI1012 or LI1102
- Information Retrieval and Source Criticism
- How2write a scientific literature review (LI1013)
- Theory of Information Retrieval
- <http://www.kth.se/en/kthb/tjanster/kurser-informationssokning/kurser-1.264017>



# Thank you! Questions?

