

Systematic development and documentation of cross-disciplinary bibliometric search queries

Mette Venås Overballe-Petersen
Special Adviser



Bibliometric analyses of cross-disciplinary research

- Search queries are often used to define cross-disciplinary research areas:
 - Not an exact science
 - Iterative process
- If large, vaguely defined and/or covering many different scientific disciplines:
 - A time-consuming approach
- Case (based on Scopus):
 - Defining and continuously developing search queries for green research by the Danish Agency for Higher Education and Science:
 - Bibliometric analysis of Denmark's green research, Sept. 2020
 - Bibliometric analysis of Denmark's space-based green research, Sept. 2022
 - Updating the bibliometric analysis of Denmark's green research, autumn 2023

A systematic way of developing, assessing and documenting the search queries

- Either from scratch or from existing search queries:
 - Brainstorming possible search terms based on:
 - Policy documents, reports, examples of research papers within the area, abstracts from Scopus, homepages of relevant organisations etc.
- Search terms are first assessed alone and then as part of a search block
- Assessment criteria
- Documentation of choices and deselected words:
 - Using Excel for the brainstorm and to document search terms
 - Compiling all choices in a Word document:
 - No. of hits
 - How precise is the search term?
 - Why chosen – not chosen?

Brainstorm overview of search terms used to update search block on sustainable food and fodder proteins

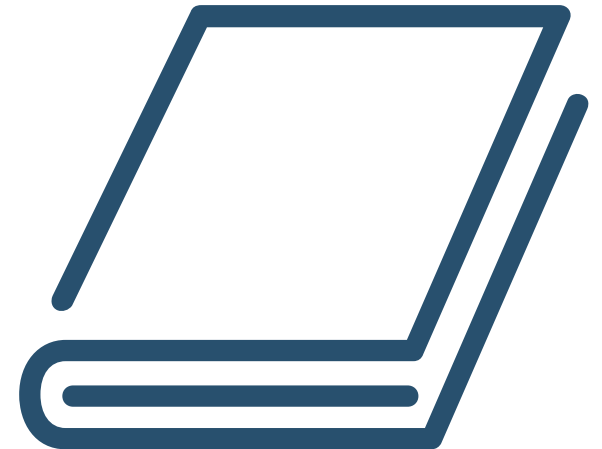
Sustainable food and fodder protein (Søgeord til opdatering af udvalgte delblokke i den eksisterende grønne søgestreng fra maj 2022 - den nye opdatering er afsluttet marts 2023)						
<u>Emneord - kombineres med w/30 med eksisterende søgeblok med bæredygtige ord</u>	<u>Opslag i Scopus</u>	<u>Emneord - kan IKKE stå alene</u>	<u>Opslag i Scopus</u>	<u>Generelle ord - kan IKKE stå alene - måske i kombinationsblok</u>	<u>Opslag i Scopus</u>	<u>Markeringer:</u>
Fungal protein	50.652	Taste and Deliciousness	1	Climate change mitigation	10.770	Markeret med grøn:
Bacterial protein	199.508	Taste-Deliciousness	0	Food security	53.735	Færdigvurderet
Seaweed protein	70	Taste & Deliciousness	0	Food & nutrition	1.509	
Grass protein	61	{Taste & Deliciousness}	0	{Food & nutrition}	145	Markeret med lysegrøn:
Feed additive	9.394	Taste	117.557	Food and nutrition	6.559	Indgår som kombinationssøgeord i delblok med grønne søgeord
From proteins	12.805	Deliciousness	113	Food nutrition	1.509	
Protein from	86.181	Accessibility and Antinutritional effects	0	Food nutrition and health	213	Markeret med lys grøn og overstreget:
Umami taste from non-animal proteins	0	Accessibility & Antinutritional effects	0	Gut microbiome	16.918	Søgeord fremsøger grøn forskning, men frasorteres i denne analyse af diverse årsager;
Non-animal protein	27	{Accessibility & Antinutritional effects}	0	Gut microbiota	42.886	f.eks. hvis det ikke kan stå alene i en enkeltbloksøgning, men godt vil kunne indgå i en
*algae protein	166	Accessibility Antinutritional effects	0	"gut microbiome" or "gut micobiota"	52.964	kombinationsblok, der dog ikke er relevant pt.
Macroalgae protein	8	Antinutritional effects	147			
Novel protein	11.641	Accessibility	226.256			Markeret med grå:
Single cell protein	1.840	Fermentation	201.811			Samlet søgning på flere søgeord
Protein	7.036.800	Biorefining	1.728			
Peptide	1.057.213	Biorefin*	16.947			Overstregede søgeord:
Amino acid	1.335.896	Sidestream	2.096			Kan ikke anvendes som grønt søgeord i de aktuelle søgeblokke
Green protein	121	Draff	67			
Protein to feed	29	Mash	7.170			
Protein ingredient	911	Umami taste	1.204			
Edible insects	1.578	Green	845.114			
Leaf protein	1.983	Green biomass	846			
Microbial protein	3.302	Green biorefin*	179			
Protein source	11.733	Protein resource	532			
Vegetable protein	42.484	Underutilized plants	233			
Alternative feed	1.177	Legume	67.939			

2.2 Alternative and sustainable food and fodder protein and ingredients/additives

2.2.a [OR]		2.2.b [OR]
Alternative feed	W/30	Biorefin*
Alternative protein		Changing climate
Bioactive peptide		Climate adapt*
Edible insects		Climate change
Feed additive		Climate friendly
Feed ingredient		Climate neutral
Feed protein		Climate smart
Food ingredient		Climate-ready
Food protein		Climatic adapt*
Grass protein		Climatic change
Green protein		Eco-friendly
Insect protein		Ecosystem functions
Leaf protein		Ecosystem services
Macroalgae protein		Environmental
Microalgae protein		Environmentally friendly
Non-animal protein		Fermentation
Novel protein		Food security
Peptide ingredient		Green biomass
Plant based protein		Green technology
Plant protein		Legume
Protein source		Microbiome
Seaweed protein		Microbiota
Single-cell protein		Natural resource
Sustainable protein		Nature based solution
Vegetable protein	Protein resource	
	Recycling	
	Renewable resource	
	Sustainability	
	Sustainable	
	Taste	
	Underutilized plants	
Search query: TITLE-ABS((block 2.2.a) W/30 (block 2.2.b)) OR AUTHKEY((block 2.2.a) AND (block 2.2.b))		

Assessment criteria

- General search – all countries, years, etc.
- Danish publications
- Last years – depending of total number of publications
- Older years
- Irrelevant subject areas
- The opposite search (AND NOT / AND)
- Evaluating title, abstract and keywords
- Checking at least 20-50 publications – sometimes more
- Repeating if deemed to be unprecise



Examples of green search blocks

1.1 Conservation

1.1.a [OR]	1.1.b [OR]
Conservation	W/15 Nature
Conserving	Ecosystem
Preservation	Habitat
Preserving	Species
Protection	Wildlife
	Landscape
	Forest
	Woodland
	Grassland
	Wetland
	Disturbance factor
	Disturbance regime
	Ecological balance
	Ecological function
	Ecological proces
	Ecological structure
	Ecological tolerance
	Resilience
	Biological diversity

Search query: TITLE-ABS((block 1.1.a) W/15 (block 1.1.b)) OR AUTHKEY((block 1.1.a) AND (block 1.1.b))

2.2 Climate, is, sea levels and the Polar areas

2.2.a [OR]	2.2.b [OR]	2.2.c [OR]
Antarctic*	AND deglaciation	AND NOT drug
Arctic	glacial change	indoor
climat*	glacier	shipping
Greenland	ice cap	
Polar	ice core	
	ice cover	
	ice sheet	
	ice velocity	
	permafrost degradation	
	permafrost thawing	
	sea ice	
	sea level	

Search query: (((TITLE-ABS (block 2.2.a) OR AUTHKEY(block 2.2.a)) AND (TITLE-ABS(block 2.2.b) OR AUTHKEY(block 2.2.b))) AND NOT (TITLE-ABS(block 2.2.c) OR AUTHKEY(block 2.2.c)))

More examples of green search blocks

1.1 Wind

1.1.a [OR]

wind energy
wind turbine
off shore wind
onshore wind
floating turbine
wind power
wind farm*
airborne wind
offshore wind
wind generator

1.1.b [OR]

AND technology
renewable energy
energy system
energy production
sustainable energy
low carbon energy
electricity
power
wind map
wind speed
wind atlas
blade
rotor

Search query: ((TITLE-ABS(block 1.1.a) OR AUTHKEY(block 1.1.a))
AND (TITLE-ABS(block 1.1.b) OR AUTHKEY(block 1.1.b)))

4.2 Carbon capture and storage in forests and wood products

4.2.a [OR]

Afforestation
Agroforestry
Forest*
Industrial roundwood
Lumber
Reforestation
Silvicultur*
Timber
Tree
Wood*

4.2.b [OR]

AND Greenhouse gas
GHG
Carbon
CO₂
Methane
CH₄

4.2.c [OR]

PRE/3 Accumulation
Balance
Budget
Capture
Content
Cycle
Cyclus
Density
Dynamics
Emissions
Fingerprint
Fixation
Flux
Footprint
Pool
Recycling
Reduction
Sequestration
Sink
Stock
Storage
Uptake

Search query: ((TITLE-ABS(block 4.2.a) OR AUTHKEY(block 4.2.a)) AND (TITLE-ABS(block 4.2.b) PRE/3 (block 4.2.c)) OR AUTHKEY((block 4.2.b) PRE/3 (block 4.2.c))))

Final search query for green research

Green search query:

- Consists of 74 green search blocks of varying lengths
- Some words are self-sufficient, others are in different combinations as AND, W/n, PRE/n or AND NOT
- Takes up 16 A4 pages.



Boolean vs. proximity operators

- Boolean operators (AND, OR, AND NOT) should be used when the subject area is specific and the words are not used frequently in other kinds of research:
 - E.g. “Antarctic*” **AND** “deglaciation”
- Proximity operators (W/n, PRE/n) should be used when the subject area is broad and maybe vaguely defined:
 - E.g. “Conservation” **W/15** “Nature”
 - Does not work in lists of keywords separated with semicolons
- Loose/approximate search “” vs. exact search {}

Textual context of search in Scopus based on the analysis of space-based green research 2011-2020

- Three searches in Scopus to identify the most correct dataset:

- Search in TITLE-ABS returns 844 publications in the period.
- **Search in TITLE-ABS+AUTHKEY returns 906 publications in the period.**
- Search in TITLE-ABS-KEY returns 1056 publications in the period.

- Conclusion:

- 840 publications appear in all three searches.
- Search in TITLE-ABS discards predominantly relevant publications.
- Search in TITLE-ABS-KEY includes app. 50 percent irrelevant publications among the additional publications included in the search.
- ***Therefore, we now only search in TITLE-ABS+AUTHKEY.***

Links

- [Definition of green research, development and innovation within the framework of the Ministry of Higher Education and Science — Uddannelses- og Forskningsministeriet \(ufm.dk\)](#)
- The report of Space-based green research with the methodology (in Danish) and the single search queries (in Excel files – mostly in English):
[Rumbaseret grøn forskning - Bibliometrisk analyse af Danmarks rumbaserede grønne forskning — Uddannelses- og Forskningsministeriet \(ufm.dk\)](#)

