

# Google über alles? Wissenschaftssuchmaschinen als alternativer Weg zu relevanten Fachpublikationen in der Landwirtschaft

Hubert Schlieber

*Präsentation 21. ÖGA-Jahrestagung  
Bozen, 5.10.2011, Forum 12*

Bundesanstalt für Agrarwirtschaft / Bibliothek  
1030 Wien, Marxergasse 2  
Tel. +43 1 87736 51 7423  
hubert.schlieber@awi.bmlfuw.gv.at  
[www.agraroekonomik.at/index.php?id=bibliothek](http://www.agraroekonomik.at/index.php?id=bibliothek)

1. Ausgangssituation: Google vs. Fach-DB
2. Alternativen: Wiss-SuMa + Open Access (OA)
3. Struktur-, Leistungsvergleich ausgewählter Wiss-SuMa
4. and the Oscar goes to .....
5. Quellen

## 1.1 Ausgangslage Frühjahr 2011 punkto SM

minimalistische Universal-SM wie Google (Suchschlitz) werden zusehends unbrauchbar für wiss. Recherchen:

- exponentiell *wachsender Datenmüll* durch unstrukturierte, unsegmentierte, unspezifische Inhalte
- *Qualität, Authentizität und Herkunft* der *Quellen* oft völlig unklar und unbestimmbar
- selten *Metadaten* vorhanden
- *Filtermöglichkeiten* VOR der Suche marginal, nicht treffsicher, NACH der Suche → *Ergebnislisten* nicht sortier- und weiterverwendbar
- *nicht nachvollziehbares Ranking* d. Ergebnisse

## 1.2 Die heile Gegenwelt am Campus / Hochschul-Community

Studenten, Lehrpersonal, Forscher, etc. an Universitäten mit UBs und IVS sowie große Forschungseinrichtungen haben es gut:

- *kostenloser* od. günstiger *Zugang* zu kosten- und lizenzpflichtigen Volltext- und Faktendatenbanken
- zielsicher und verlässlich *relevante* und *validierte* Information darin auffindbar
- *Beispiele*: Scopus, Web of Science, SpringerLink, SCI, CAB, CSA, ...

## 1.3 Is there anybody out there ....?

aber sicher:

- *kleine* oder kaputtgesparte *Forschungseinrichtungen* (ohne Lobby), auch Bundesanstalten (BMLFUW), kleinere Bibliotheken, Schulen ...?
- *Privatforscher*, interessierte Laien, Lehrer, Schüler
- is there a „Hope of Deliverance“?

einen gangbaren Ausweg und brauchbare Alternativen bieten

### Spezial- oder Wissenschafts-Suchmaschinen (Wiss-SM)

- ❖ zielen weltweit auf Inhalte der div. Open-Access (OA)<sup>1</sup> Plattformen (Dokumentenserver, Repositories) ab
- ❖ erreichen Inhalte des *Deep Web*<sup>2</sup> (Invisible Web): bis zu 500x größer als das Visible Web (Google u. Konsorten)

1 freier und kostenloser Zugang zu und Gebrauch von wissenschaftlicher Literatur und anderen Materialien im Internet

2 Webinhalte, die aus technischen oder organisatorischen Gründen nicht für Universalsuchmaschinen durchsuchbar sind

Was machen nun Wiss-SuMa besser als der Marktführer Google und Konsorten?

1. Abwendung vom Universalitätsanspruch hin zu *bewusst ausgewählten* fachlich relevanten, wiss. *Quellen* weltweit; dies wird zumeist *transparent* und *explizit* gemacht
2. Daten weisen nötige *Metadaten* auf (Autor, Datum, Keywords, Sprache, Fundstelle, Datenformat, etc.)
3. umfangreiche *Filter* VOR und NACH der Suche: für WissenschaftlerInnen gilt es, nicht soviel wie möglich aufzulisten, sondern die Datenflut auf *Relevantes* so eng wie möglich zu *beschränken* (Gleichgewicht von „Precision“ and „Recall“)<sup>1</sup>

<sup>1</sup> **Recall, Precision** ... sind Maße zur Beurteilung der Güte von Treffermengen einer Recherche beim Information Retrieval... Der *Recall* gibt mit dem Anteil der bei einer Suche gefundenen relevanten Dokumente die **Vollständigkeit** eines Suchergebnisses an. Die *Precision* beschreibt mit dem Anteil relevanter Dokumente an der Ergebnismenge die **Genauigkeit** eines Suchergebnisses...”

4. Ergebnisse können vielfältig *sortiert, kopiert, exportiert* oder *gespeichert* werden
5. Auswahl der gewünschten *Publikationsart* möglich (Fachartikel, Monographie, Forschungs-, Tagungsbericht, statistisches Material, Hochschulschrift, e-journal, pre-print, Bildmaterial,.....)
6. Einengung der Suche auf hochwertige, *kostenlose* Information möglich (Open Access)
7. *redaktionell* gepflegte + gewartete Daten verfügbar mit genormter Schnittstelle ([OAI-Protokoll](#))



### 3.1 Ausgangslage

- zur besseren *Vergleichbarkeit* wurden bei jeder Wiss-SuMa dieselben Themen bzw. Suchbegriffe in englisch + deutsch abgefragt
- *aktuelle* und *brisante Themen*: Biologische Landwirtschaft, LW und Klimawandel, GAP-Reform, Risikomanagement, Nachhaltigkeit, Gentechnik, Agrarförderung, Nachwachsende Rohstoffe (Nawaro), Bioenergie, Biomasse
- *Themenvorschläge* stammen aus Kollegen/-innenkreis

## 3.2 Drei Typen von SM unterschieden

1. SM *speziell* für Open-Access-Angebote ([BASE](#), [MS Academic Search](#), [Scientific Commons](#), [Cite Seer](#), [Open-J-Gate](#), [OpenDOAR](#), [ROAR](#), [OAISTER](#), [DOAJ](#), [SpringerLink OA](#))
2. SM für OA und kommerzielle Angebote *vermischt* ([Scirus](#), [Google Scholar](#))
3. *Repositories* mit Suchoberflächen ([ScienceGate](#), [AgEcon](#), [Organic eprints](#), [eDOC](#))

### 3.3 Check der SM auf Vorhandensein von

1. *Ergebniserweiternden* Merkmalen bzw. Suchfeatures (=mehr, aber weitläufigere Themen): Volltextsuche, Stemming<sup>1</sup>, themengeführte Suche, Trunkierung, logisches ODER
2. *Ergebniseinschränkenden* Features (=weniger, aber präzisere Treffer): Suche nach Metadaten (Autor, Jahr, Keyword,...), Filter und Sortieroptionen vor und nach Suche, logisches UND
3. *Sonstige*: Quellenangabe, Anzahl aller vorhandenen Medien, RSS-Abo, Suchhistorie nachvollziehbar, Export von Suchergebnissen, Trefferaktualität

<sup>1</sup> (Grundformenreduktion, Normalformenreduktion)

## 3.4 Vorgangsweise und Durchführung

- Suchstring, Trefferzahl, sonstige Features/SM als *Matrix* in MS-Excel aufgetragen → praktisch für Mehrfachauswertung
- *Unterscheidung* von themenabhängiger (Treffer-) Leistung (s. 3.1) bzw. themenunabhängiger Leistung (s. 3.3)
- *Gruppierung* nach 3 verschiedenen SM-Typen (s. 3.2)
- *Gewichtung* der Ergebnisse nach Themen(un)abhängigkeit
- Als *Beispiel* zwei Screenshots eines Excel-Arbeitsblattes

# 3 Struktur-, Leistungsvergleich ausgewählter Wiss-SuMa

Open\_Access\_Landwirtschaft32.xls [Kompatibilitätsmodus] - Microsoft Excel

1	SM (Open Access)																					
2	Suchmaschine	BASE		MS Academic Search		Scientific Commons		CiteSeer ≡		Open-J-Gate		OpenDOAR ≡		ROAR ≡		DOAJ		OAISTER		Springer Link (OA) ≡		
3	Kürzel	BA		MSA		SCO		CS		OJ		OD		RO		DO		OAI		SL		
4		Gesamt	Keyword	Gesamt	Keyword	Gesamt	Keyword	Gesamt	Keyword	Gesamt	Keyword	Gesamt	Keyword	Gesamt	Keyword	Gesamt	Keyword	Gesamt	Keyword	Gesamt	Keyword	
5	<i>Biolandbau dt.</i>	2.668	2.609		x	408	x					191	x		x			428	32	71	x	
6	<i>Biolandbau en.</i>	10.112	835	1.178	x	8.344	x	339	4	431	261	829	x	649	x	151	68	3.424	791	2.047	x	
7	<i>Klimawandel dt.</i>	137	8		x		x					347	x	160	x			24	9	194	x	
8	<i>Klimawandel en.</i>	2.789	268	821	x	4.093	x	3.474	3	183	28	671	x	486	x	105	8	1.676	157	10.000	x	
9	<i>GAP dt.</i>	117	25		x		x					123	x		x			19		29	x	
10	<i>GAP en.</i>	3.286	243	343	x	2.448	x	248	2	42	11	521	x	398	x		18	1.728	119	653	x	
11	<i>Risiko dt.</i>	89	3		x	122	x					390	x	230	x					574	x	
12	<i>Risiko en.</i>	1.050	96	133	x	1.698	x	1.252				540	x	282	x			807	277	1.522	x	
13	<i>Nachhaltigkeit dt.</i>	1.448	42		x		x					377	x	220	x			304	27	955	x	
14	<i>Nachhaltigkeit en.</i>	14.815	4.297	2.097	x	5.576	x	10.262	7	1.064	203	704	x	505	x	275	60	10.376	1.795	22.823	x	
15	<i>Gentechnik dt.</i>	189	6		x	110	x					242	x	94	x					218	x	
16	<i>Gentechnik en.</i>	975	363	136	x	2.240	x	652			13	624	x	458	x			908	204	4.292	x	
17	<i>Förderung dt.</i>	104	97		x		x					99	x		x					9	x	
18	<i>Förderung en.</i>	7.614	974	463	x	4.044	x	2.534			18	594	x	356	x	44		3.568	296	4.535	x	
19	<i>NAWARO dt.</i>	3.824	3.748	252	x	2.035	x	415	6			601	x	324	x			3.322	1.145	1.071	x	
20	<i>NAWARO en.</i>	51.763	50.153	19.787	x	62.898	x	6.850	69	7.171	1.668	1.000	x	688	x	3.096	463	43.652	15.030	64.387	x	
21		#####	63.767	25.210		94.016		26.026	91	8.891	2.202	7.853		4.850		3.671	617	70.236	19.882	#####	##	
22	Besondere Eignung der Suchmaschine für Thema (Rangordnung)	Nawaro,en 4 Nachhalt,en 3 BL,en 2 Förd,en 1		Nawaro,en Nachhalt,en BL,en Klima,en		Nawaro,en BL,en Klima,en Förd,en		Nachhalt,en Nawaro,en Klima,en Förd,en		Nawaro,en Nachhalt,en BL,en Klima,en		Nawaro,en BL,en Nachhalt,en Klima,en		Nawaro,en BL,en Nachhalt,en Klima,en		Nawaro,en Nachhalt,en BL,en Klima,en		Nawaro,en Nachhalt,en BL,en Förd,en		Nawaro,en Nachhalt,en Klima,en Förd,en		
23																						
24																						
25																						
26																						
27		BASE		MS Academic Search		Scientific Commons		CiteSeer ≡		Open-J-Gate		OpenDOAR ≡		ROAR ≡		DOAJ		OAISTER		Springer Link (OA) ≡		

# 3 Struktur-, Leistungsvergleich ausgewählter Wiss-SuMa

Open\_Access\_Landwirtschaft32.xls [Kompatibilitätsmodus] - Microsoft Excel

	B	C	D	E	F	G	H	I	J	K	L	M	N	O	P	Q	R	S	T	U	V
24																					
25																					
26																					
27		BASE	MS Academic Search	Scientific Commons	CiteSeer ≡	Open-J-Gate	OpenDOAR ≡	ROAR ≡	DOAJ	OAISTER	Springer Link (OA) ≡										
28	Volltextsuche	0,5	0	0	1	0	1	1	0	0	1										
29	Stemming	1	0	0	0	0	1	1	0	0	0										
30	Thesaurus	1	0	0	0	0	0	0	0	0	0										
31	Themenfelder	0	1	0	0	1	1	1	0	0	0										
32	Maskierung	1	0	0	1	0	0	0	0	0	0										
33	Boolsche Verkn. ODER	1	0	1	1	1	1	1	1	1	1										
34	Exakte Suche	1	1	0	0	1	1	1	1	1	1									0	
35	Keywordsuche	1	0	0	1	1	0	0	1	0	0										
36	Suche in sonst. Elementen	1	1	1	1	1	1	1	1	1	1										
37	Filter VOR Suche	1	1	0	1	1	1	1	1	1	0										
38	Filter/Sort. NACH Suche	1	1	1	1	0	1	1	0	1	0										
39	Relevanz-Ranking	1	1	1	1	0	1	1	0	1	0										
40	Boolsche Verkn. UND	1	1	1	1	1	1	1	1	1	1										
41	Quellenangabe	1	0	1	0	1	1	1	1	1	1										
42	Angabe Anzahl DE	1	1	1	0	0	0	1	0	0	1										
43	RSS	0	1	1	1	0	1	0	0	0	0										
44	Speichern/Export	0	0	1	1	1	0	0	0	0	0										
45	Suchhistorie	1	0	0	0	0	0	0	0	0	0										
46	Web 2.0	0	1	0	1	0	0	0	0	0	0										
47		14,5	10	9	12	9	12	12	7	14	11										1
48		BASE	MS Academic Search	Scientific Commons	CiteSeer ≡	Open-J-Gate	OpenDOAR ≡	ROAR ≡	DOAJ	OAISTER	Springer Link (OA) ≡										
49																					
50																					

- ✓ *Sieger bei Trefferanzahl (Recall):* **SCIRUS** (> 6 Mio. Titel), gefolgt von GOOGLE SCHOLAR und BASE, wobei letztere Volltextsuche erlauben, was der Präzision des Ergebnisses nicht immer zuträglich ist. Trend ersichtlich → Titel zu *Nawaros* wurden praktisch überall nachgewiesen, gefolgt von Nachhaltigkeit, Biolandbau und Klimawandel
- ✓ *Sieger bei erweiterten Suchfeatures (Precision):* **BASE** vor OAISTER und SCIRUS, wobei die Bool'schen Operationen UND/ODER sowie die Suche in Metadatenfeldern bei allen Wiss-SuMa aufzufinden waren.

Hermann, L. (2009): Möglichkeiten und Grenzen von Suchmaschinen bei der wissenschaftlichen Recherche im Internet : Masterarbeit. - Berlin : Humboldt - Univ. Berlin. Online:

<http://edoc.hu-berlin.de/master/hermann-lars-2010-01-28/PDF/hermann.pdf>

Lewandowski, D. (2007a): Deep Web - Suchen wir was man finden könnte? Online:

[http://www.durchdenken.de/lewandowski/doc/Euroforum\\_Lewandowski.pdf](http://www.durchdenken.de/lewandowski/doc/Euroforum_Lewandowski.pdf)

Lewandowski, D. (2007b): Google, Deep Web und Fachdatenbanken. Online: <http://www.durchdenken.de/lewandowski/doc/swp2007.pdf>



Müller, D. (2011): Das Web versinkt im Datenmüll. Blogbeitrag.-  
<http://tinyurl.com/6dz9992>

Schlieber, H. (2008: Der freie (kostenlose) Zugang zu Publikationen aus Wissenschaft und Bildung im Internet über Open Access (OA) : Schwerpunkt Landwirtschaft ; August 2008. - Wien : AWI  
[http://eprints.rclis.org/14781/1/Open\\_Access.pdf](http://eprints.rclis.org/14781/1/Open_Access.pdf)

Strzolka, R. (2008): Das Internet als Weltbibliothek: Suchmaschinen und ihre Bedeutung für den Wissenserwerb. - Berlin : Simon Verl. f. Bibliothekswissen

- Danke für ihre Aufmerksamkeit, und –
- bitte selbst ausprobieren, es macht Spaß

Happy Searching