

TREND MINING IRIS: 20 JAHRE ENTWICKLUNG DER RECHTSINFORMATIK

Bettina Mielke / Christian Wolff

Dr. iur., Vorsitzende Richterin am Landgericht Regensburg, Lehrbeauftragte an der Universität Regensburg
Kumpfmühler Straße 4, 93047 Regensburg, DE
bettina.mielke@lg-r.bayern.de

Professor, Institut für Information und Medien, Sprache und Kultur, Lehrstuhl für Medieninformatik
Universität Regensburg, 93040 Regensburg, DE
christian.wolff@ur.de, <http://mi.ur.de>

Schlagnote: *Trendanalyse, Text Mining, Corpuslinguistik, Entwicklung der Rechtsinformatik*

Abstract: *Rechtsinformatik ist ein junges und interdisziplinär geprägtes wissenschaftliches Arbeitsgebiet. Anlässlich des 20-jährigen Jubiläums des internationalen Rechtsinformatik-Symposiums Salzburg (IRIS) untersuchen wir mit verschiedenen Methoden, darunter auch corpuslinguistischen, die Entwicklung dieses Faches. Wir betrachten dabei die Tagungsbände der einzelnen Jahrgänge als Teilcorpora, die jeweils für einen bestimmten Zeitpunkt den aktuellen Entwicklungsstand repräsentieren. Dadurch lassen sich Zeitscheibencorpora gewinnen, wobei jeder Tagungsband eine Zeitscheibe darstellt. Diese Zeitscheiben werden im Vergleich zueinander analysiert, um über den Zeitraum von 20 Jahren Entwicklungstrends aufzuzeigen. Dabei geht es um inhaltliche Analysen, die auf dem Text der einzelnen Bände aufsetzen.*

1. Einführung

In unserem Beitrag wollen wir untersuchen, ob sich aus 20 Jahren Internationales Rechtsinformatik-Symposium Salzburg (IRIS) Trends der Rechtsinformatik ablesen lassen. Dafür analysieren wir die seither erschienenen Tagungsbände von 2000 bis 2016 (in den Jahren 1998 und 1999 sind keine Proceedings erschienen) inhaltlich. Wir wollen dabei über bibliometrische bzw. szientometrische Untersuchungen, die eher quantitative Analyseaspekte in den Vordergrund rücken (etwa Zitationsraten, also die Häufigkeit, mit der bestimmte Autoren zitiert werden¹), hinausgehen. Dazu dienen uns corpusbasierte Analyseverfahren, die sich in den vergangenen zwei Jahrzehnten parallel zur Entwicklung des IRIS etabliert haben.

Zu fragen ist danach, was Indikatoren und Bausteine einer solchen Trendanalyse sein können. Naheliegend sind Begriffe und ihre Verwendungshäufigkeiten, da dadurch dargestellt werden kann, welches Fachvokabular ein Fach verwendet und welche Begriffe sich im zeitlichen Verlauf wie entwickeln, also z.B. häufiger oder weniger häufig gebraucht werden. Der hier betrachtete Anwendungsfall soll dazu beitragen, die Untersuchung solcher Prozesse stärker zu systematisieren. Daran besteht offenkundig Interesse, wie vergleichbare Analysen im Bereich der Wirtschaftswissenschaften, der Psychologie und der Szientometrie zeigen².

Die Analyse von Trends ist nicht neu, vielmehr gehört das Aufzeigen von Entwicklungstendenzen im Zeitverlauf zu den typischen Erkenntnisinteressen in vielen Disziplinen. Die konkrete Analyse auf der Basis digitaler Dokumentrepräsentation, d.h. umfangreicher Volltextkollektionen, dürfte aber eher noch Seltenheitswert besitzen. Für die Trendanalyse sind verschiedene Teildisziplinen von Bedeutung:

¹ Vgl. etwa BALL 2014.

² SIDOROVA/EVANGELOPOULOS/VALACICH/RAMAKRISHNAN 2008 und ROONEY/McKENNA/BARKER 2011 (Wirtschaft), CRETCHLEY/ROONEY/GALLOIS 2010 (Psychologie) und RAVIKUMAR/AGRAHARI/SINGH 2015 (Szientometrie).

1. Angewandte Informatik, Corpuslinguistik, Sprachtechnologie

Voraussetzung für eine angemessene Verarbeitung großer Dokumentenmengen sind entsprechende Werkzeuge, die sowohl aus der Sprachtechnologie als auch aus der angewandten Informatik stammen können. Hier liegen mittlerweile viele und z.T. online verfügbare Werkzeuge vor. Das Spektrum solcher Werkzeuge reicht von Suchmaschinentechnologie, die große Volltextmengen durch Indexierung zugänglich machen kann, bis hin zu stärker auf corpuslinguistischen Fragestellungen ausgerichteten Anwendungen.

2. Informationswissenschaft, Bibliometrie bzw. Szientometrie

Vorliegend geht es um die Analyse der Entwicklung einer bestimmten Fachdisziplin – der Rechtsinformatik – in einer bestimmten Region (primär der zentraleuropäische Raum mit Schwerpunkt auf den deutschsprachigen Ländern). Die Untersuchung einer Wissenschaftsdisziplin im Zeitverlauf kann man dem Fächerspektrum der Bibliotheks- und Informationswissenschaft und konkret den fachlichen Schwerpunkten der Biblio- bzw. Szientometrie zurechnen. Biblio- und Szientometrie sind allerdings seit ihrer Einführung in den 1950er Jahren stark durch rein quantitative und standardisierte Auswertungsmechanismen geprägt gewesen, die (noch) nicht den Inhalt (Volltext) der untersuchten Literatur auswerten. Dazu gehören etwa die Berechnung von Zitationsindizes mit Hilfe von Datenbanken wie dem *science citation index*, die Anwendung von Metriken wie dem Hirsch-Index oder die Analyse von Autorenetzwerken, also Verfahren, die nicht die durch Text repräsentierten Inhalte analysieren.³ Eine fach- und inhaltsbezogene Auswertung gewinnt hingegen mit der Verfügbarkeit digitaler Volltexte und passender Werkzeuge an Bedeutung.

3. Rechtsinformatik

Das betrachtete Fach selbst, also die Rechtsinformatik, liefert Hinweise auf die zu untersuchenden Fragen. Hier geht es insbesondere darum, aus großen Analysemenen die richtigen Begriffe, Personen oder Konzepte auszuwählen, deren tatsächliche Entwicklung näher zu betrachten als besonders lohnenswert erscheint.

2. Forschungsfragen

Die Untersuchung der Entwicklung des IRIS in den letzten 20 Jahren, genauer der bislang erschienenen Tagungsbände (von 2000 bis einschließlich 2016) wirft folgende Forschungsfragen auf:

1. Was sind die grundlegenden quantitativen Eckdaten des IRIS (Anzahl der Artikel, Seitenumfang etc.)?
2. Welche Schwerpunkte lassen sich für die verschiedenen Jahre ausmachen?
3. Welche Schwerpunkte sind im Zeitverlauf neu entstanden, schwächer geworden oder weggefallen?
4. Schlagen sich die Themenschwerpunkte der verschiedenen Jahre erkennbar in den Beiträgen nieder?
5. Haben sich die diskutierten Konzepte selbst verändert, m.a.W.: hat sich das semantische Umfeld von Schlüsselbegriffen der Rechtsinformatik im betrachteten Zeitraum verändert, und wenn ja, wie?
6. Welche Akteure sind beteiligt und wie haben sich die Beziehungen zwischen den Akteuren verändert?

Nicht alle Fragen können hier geklärt werden, im Folgenden konzentrieren wir uns auf die Fragen 1 bis 4.

3. Methodik

Wir setzen für diesen Beitrag sowohl die intellektuelle Analyse von Teilstrukturen der IRIS-Bände als auch werkzeuggestützte Verfahren des Text Mining ein, ersteres für die Analyse der Themenschwerpunkte, letzteres für die Untersuchung des sprachlichen Materials auf der Basis der Volltexte der Tagungsbände.

Wie bereits in drei vorangegangenen IRIS-Beiträgen⁴ haben wir uns auch bei dem hier vorliegenden Vorhaben für die Nutzung der *Voyant Tools*⁵ als Analysewerkzeuge entschieden. Dafür sprach einerseits die bereits im

³ GARFIELD 1955; BALL 2014, S. 53, 77.

⁴ MIELKE/WOLFF 2013, 2015 und 2016.

⁵ SINCLAIR/ROCKWELL 2016.

Umgang mit diesen Werkzeugen gewonnene Erfahrung, andererseits auch die seit 2015 erfolgte nicht unerhebliche Weiterentwicklung der Werkzeuge, die mittlerweile in einer funktional deutlich erweiterten Fassung vorliegen. Die Werkzeuge, die von STÉFAN SINCLAIR (McGill University) und GEOFFREY ROCKWELL (University of Alberta), zwei kanadischen Wissenschaftlern auf dem Gebiet der Digital Humanities entwickelt wurden⁶, sind online zugänglich⁷ und stellen die erforderliche Verarbeitungs- und Speicherungsinfrastruktur für die Verarbeitung auch größerer Textmengen zur Verfügung. Ein Nachteil beim Einsatz dieser Werkzeuge liegt in der lückenhaften Dokumentation und der deshalb nicht immer gegebenen Nachvollziehbarkeit der Arbeitsweise der Werkzeuge und ihrer Analyseergebnisse.

Die Tagungsbände des IRIS 2000 bis 2016 wurden durch die Herausgeber bzw. den Verlag bereitgestellt. In 15 von insgesamt 17 Fällen lag entweder die Druckvorlage als PDF oder eine Vorstufe dazu vor, zum Beispiel einzelne Word-Dateien der jeweiligen Beiträge. Für die Jahrgänge 2012 und 2014 war es erforderlich, die online als Ausgaben des *Jusletter IT*⁸ verfügbaren Artikel herunterzuladen und in ein entsprechendes Format zu konvertieren. Das Zielformat für die weitere Verarbeitung des Textkorpus war dabei PDF. Zum Teil war eine weitgehendere Bereinigung bzw. Datenkonvertierung erforderlich, etwa dort, wo außerhalb der üblichen Kodierungsbereiche Glyphen zur Schriftformatierung eingesetzt worden waren (etwa Kapitalchen im IRIS-Band 2016). Da der primäre Analysefokus auf der Entwicklung des IRIS insgesamt liegt, erschien es ausreichend, ein Corpus aufzubauen, dessen einzelne Dokumente jeweils einen vollständigen Tagungsband umfassen. Für zukünftige Analysen wäre dagegen auch eine kleinteiligere Modellierung auf Artikelebene oder sogar darunter bis hin zu den Strukturelementen der einzelnen Artikel (Überschriftentexte, Abbildungen, Tabellen, Literaturverzeichnis etc.) denkbar und wünschenswert. Insofern liegt nun ein Corpus vor, das aus 17 IRIS-Tagungsbänden besteht. Sein Gesamtumfang liegt bei knapp 10'000 Druckseiten sowie deutlich über 1'000 Artikeln. Diese Datenbestände wurden mit den *Voyant Tools* analysiert. Als weiterer Bereinigungs-schritt wurde eine deutsch-englische Stoppwortliste zur Filterung auf das Corpus angewandt, die auf einer Datensammlung aus vorangegangenen Untersuchungen beruht⁹ und die für die Analyse des IRIS geringfügig modifiziert wurde. Der Einsatz einer Stoppwortliste ist für häufigkeitsorientierte Betrachtungen sinnvoll, da sonst ausschließlich nicht inhaltstragende Stoppwörter die Ranglisten dominieren. Mit der Verarbeitung in *Voyant Tools* steht für die Betrachtung des Corpus die ganze Palette der dort angebotenen Werkzeuge zur Verfügung. Ausgewählte Analysen bzw. Visualisierungsformen sollen dies im nachfolgenden Ergebnisteil illustrieren.

4. Die IRIS-Tagungsbände 2000–2016 im Überblick

Um einen Überblick zur Entwicklung der wichtigsten Themenfelder des IRIS zu gewinnen, haben wir die Inhaltsverzeichnisse und insbesondere die Struktur der Tagungsbände näher betrachtet, in denen jeweils mehrere Artikel einer Themengruppe zugeordnet sind.

Hinsichtlich des Gesamtumfangs der Tagungsbände ergibt sich eine Entwicklung von 230 Seiten im Jahr 2000 bis zuletzt regelmäßig deutlich über 600 Seiten. Dabei umfasst bereits der zweite Tagungsband aus dem Jahr 2001 424 Seiten. Zwischen 2002 und 2009 betragen die Seitenzahlen von 378 Seiten (2009) bis 662 Seiten (2005). Ab 2010 ist der Umfang in etwa gleichbleibend zwischen 618 Seiten (2011) und 692 Seiten im Jahr 2014. Nicht berücksichtigt werden konnte der Umstand, dass der Umfang einer Seite insbesondere je nach Verlag (Verlag Österreich, Boorberg, books@ocg.at) schwankt. Auch die Zahl der Artikel steigt von 25 Artikeln im ersten Band bis zu 129 Beiträgen im Jahr 2012, in dem sich allerdings auch zahlreiche Abstracts als eigene Beiträge befinden. Abbildung 1 visualisiert mittels Liniendiagrammen die Entwicklung des IRIS in quantitativer Hinsicht (Seitenzahlen und Anzahl von Artikeln).

⁶ Vgl. MIELKE/WOLFF 2013, S. 378.

⁷ <https://voyant-tools.org/> (alle Websites geprüft am 26. Januar 2017).

⁸ <http://jusletter-it.eu/>.

⁹ Vgl. MIELKE/WOLFF 2016, S. 134.

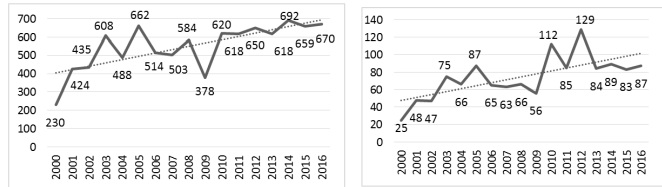


Abbildung 1: Seitenumfang (links) bzw. Anzahl der Artikel (rechts) in den IRIS-Bänden

5. Entwicklung der Themengruppen der IRIS-Bände 2000–2016

In einem weiteren Schritt wurden alle in den 17 Tagungsbänden genannten Themengruppen erfasst und die jeweils von den Herausgebern der Tagungsbände diesen Themengruppen zugeordneten Artikel gezählt. Soweit es sich lediglich um Schreibvarianten desselben Konzeptes handelte (zum Beispiel Electronic Commerce, ECommerce, e-Commerce, E-Commerce), haben wir die Varianten sofort zusammengeführt. Nach diesem ersten Analyseschritt ergaben sich insgesamt über 60 verschiedene Themengruppen in den IRIS-Bänden seit 2000. Da zahlreiche Themengruppen einen starken inneren Zusammenhang aufweisen, konnte diese Liste weiter verdichtet werden. So wurde etwa die in den Anfangsjahren – ein offensichtliches Zeitphänomen – noch klar erkennbare Themengruppe *elektronisches Publizieren* dem Bereich der Rechtsinformation zugeordnet. Dies erfolgt auf der Basis der Überlegung, dass einer der traditionellen Themengruppen des IRIS die juristische Informationswissenschaft bzw. Fachinformation für die Rechtswissenschaft ist und dass hier ein offensichtlicher Zusammenhang zwischen dem Anbieten von Dienstleistungen der Rechtsinformation, dem Aufbau und der Entwicklung von Rechtsdatenbanken und dem elektronischen Publizieren besteht. Man darf davon ausgehen, dass das elektronische Publizieren mittlerweile zur Selbstverständlichkeit geworden ist, sodass es als Thema nicht mehr explizit genannt wird. Auch in anderen Fällen war eine Zusammenführung einzelner Themengruppen sinnvoll und möglich. Soweit Themengruppen in einigen Bänden zusammengefasst waren (z.B. *e-Commerce* und *e-Learning* in den Bänden 2007 und 2008 oder *E-Government* und *E-Justiz* im Band 2010), wurden die Aufsätze aufgrund ihres Inhalts den einzelnen Themengruppen zugeordnet. Eines der Ziele war dabei, die wichtigsten Schwerpunkte des IRIS herauszuarbeiten und zu untersuchen, ob sich im Zeitverlauf auffällige Entwicklungen nachweisen lassen. Bei dieser Methodik ist zu beachten, dass die Themengruppen maßgeblich durch die jeweiligen Herausgeber bestimmt werden. Insofern wird also damit auch nachvollzogen, wie die Schlüsselakteure des IRIS dieses prägen – und sicher gleichzeitig auch von der Gesamtheit der Autorenschaft des IRIS geprägt werden. Wir gehen davon aus, dass es sich hier um einen Kreislauf aus Themensetzung, eingereichten Beiträgen und Modifikation von Themensetzungen handelt.

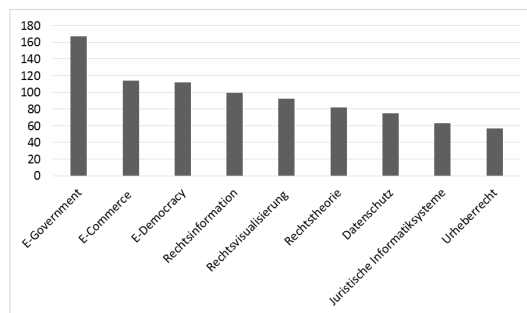


Abbildung 2: Anzahl der Artikel zu den Themen mit mehr als 50 Artikeln insgesamt

Abbildung 2 zeigt die Themengruppen, zu denen in den IRIS-Bänden insgesamt mehr als 50 Artikel publiziert wurden. Zum Thema *E-Government* finden sich mehr als 160 Artikel in 20 Jahren IRIS, am zweit- und dritthäufigsten erschienen Beiträge zu *E-Commerce* und *E-Democracy* mit jeweils über 110 Artikeln, gefolgt von den Themengruppen *Rechtsinformation*, *Rechtsvisualisierung* und *Rechtstheorie* mit Artikelzahlen von 80 bis 100, wobei die Themengruppe *Rechtsinformation* eng mit der Themengruppe *Juristische Informatiksysteme* zusammenhängt, die über 60 Beiträge enthält. Das Thema IT-Recht untergliedert sich in verschiedene Themengruppen wie etwa *Datenschutz* und *Urheberrecht* mit über 70 bzw. knapp 60 Artikeln.

Abbildung 3 zeigt, dass sich die Themengruppen im Laufe von 20 Jahren ausdifferenziert haben. Im ersten Band gab es sechs Themengruppen, im Tagungsband 2012 waren es 22, später gingen sie wieder etwas zurück.

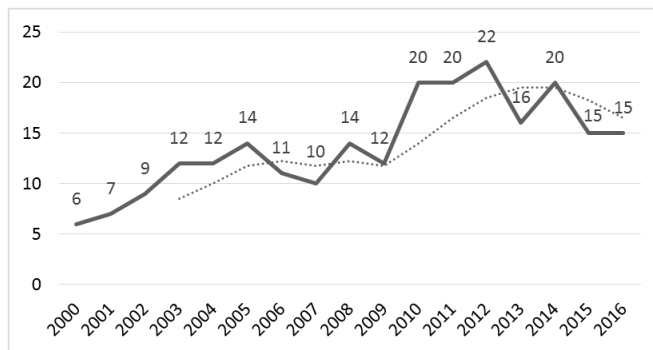


Abbildung 3: Anzahl der Themenfelder im Zeitverlauf (gestrichelte Trendlinie: Gleitender Durchschnitt über vier Jahre)

In einem weiteren Schritt zeigen wir in den nachfolgenden Diagrammen (Abbildung 4 auf der folgenden Seite) auf, wie sich der prozentuale Anteil der Artikel zu einer Themengruppe an allen Artikeln der jeweiligen IRIS-Bände seit dem ersten Tagungsband entwickelt hat. Es zeigt sich zum Beispiel, dass der Anteil der Artikel der Themengruppe *E-Government* von einem Anteil von deutlich über 20 % auf nur noch etwas über 10 % sinkt, auch die verwandte Themengruppe *Verwaltungsnetzwerke* nimmt keinen so großen Anteil ein, dass sich die Abnahme hinsichtlich des Themas *E-Government* durch eine Zunahme bei diesem Thema erklären könnte. Eine eher abnehmende Tendenz ergibt sich auch für die Themengruppe *E-Commerce*, während die Themen *Rechtsinformation* und *juristische Informatiksysteme* gemessen an der Zahl der Artikel dazu relativ stabil bleiben, insbesondere wenn man sie in Kombination betrachtet. Ähnlich ist es bei den Themen *E-Democracy* und *Rechtstheorie*. Die Zahl der Artikel zur Themengruppe *Rechtsvisualisierung* nimmt hingegen zu.

6. Inhaltliche Entwicklung – Vokabularanalyse

Im Folgenden stellen wir die Vokabularentwicklung des IRIS als Ergebnis der Auswertung der Volltexte vor.

6.1. Umfang des sprachlichen Materials (Tokens, Types)

Aus Tabelle 1 (folgende Seite) ergibt sich, dass die Zahl der *Token* (laufende Wortformen) stark zunimmt und sich von 2000 bis 2016 verfünffacht. Die Zahl der *Types* (unterschiedliche Wortformen) erhöht sich nicht im gleichen Maße, sondern verdreifacht sich lediglich, so dass die *Type-Token-Relation* deutlich sinkt. Der Umfang der Bände nimmt also schneller zu als ihre sprachliche Vielfalt. Dies bedeutet, dass man mehr Text lesen muss, um auf ein neues Wort zu stoßen, also letztlich mehr über dasselbe geschrieben wird.

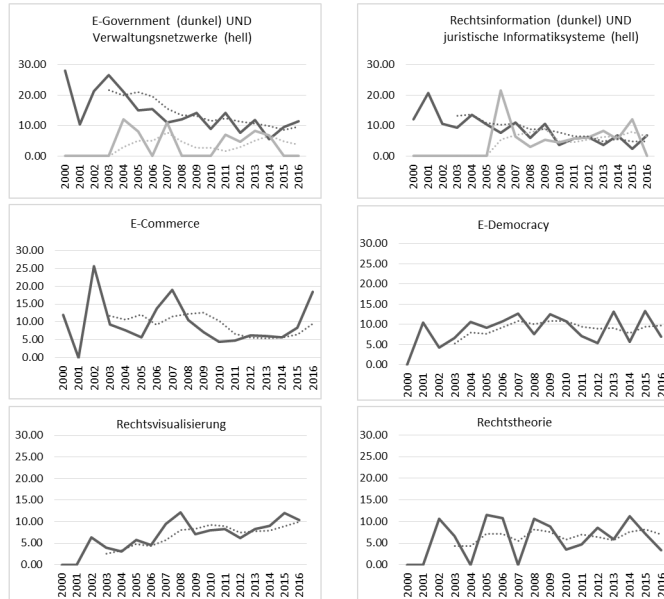


Abbildung 4: Prozentualer Anteil der Themen im Zeitverlauf (gestrichelt: Gleitender Schnitt über vier Jahre)

Jahr	Tokens	Types	TTR	Jahr	Tokens	Types	TTR	Jahr	Tokens	Types	TTR
2000	57.145	12.687	0,22	2006	140.458	30.640	0,22	2016	317.348	37.131	0,12
2001	106.246	23.585	0,22	2007	128.616	26.979	0,21	2012	266.023	31.374	0,12
2002	113.090	19.760	0,17	2008	152.351	30.469	0,20	2013	265.876	34.873	0,13
2003	177.788	28.349	0,16	2009	139.382	23.428	0,17	2014	307.495	33.883	0,11
2004	128.535	28.676	0,22	2010	243.893	32.588	0,13	2015	295.875	36.774	0,12
2005	188.665	30.213	0,16	2011	261.635	36.816	0,14				

Tabelle 1: Anzahl laufende Wortformen (Tokens), unterschiedliche Wörter (Types) sowie die Type-Token-Relation (TTR) als Maß der Vokabulardichte

6.2. Häufigste und auffälligste Begriffe

Tabelle 2 und 3 machen deutlich, dass *Recht* als einziger Begriff in jedem Jahr unter den zehn häufigsten inhaltstragenden Begriffen ist. Bei den anderen Begriffen zeigt sich ein Wandel. So nehmen zuletzt englische Begriffe zu. Gleichzeitig geht die Häufigkeit der Begriffe zurück, die auf Österreich hindeuten. So ist der Ortsname *Wien* in den Jahren 2000, 2003, 2004, 2009 und letztmals 2011 unter den *Top Ten*.

Die Begriffe *law*, *legal* und *data* sind ab 2011 immer unter den *Top Five* der inhaltstragenden Begriffe. Wenn man über die jeweils zehn häufigsten Begriffe hinausgeht und die häufigsten 300 Begriffe auswertet, sieht man, dass der Begriff *data* bis 2007 nur zweimal unter den ersten 300 Begriffen auftaucht, nämlich 2003 auf Rang 26 und 2006 auf Rang 158. 2008 ist er dann auf Rang zwölf und 2010 auf Rang 26, während er 2009 nicht unter den ersten 300 Begriffen zu finden ist. In den ersten Jahren dominieren auch die Begriffe *elektronisch* und *Internet*, die dann seltener werden. Der Begriff *elektronisch* ist bis 2006 jeweils unter den *Top Three*, ab 2007 ist er nur noch vereinzelt unter den *Top Ten* (2008 auf Rang zehn, 2010 auf Rang neun und ab 2010 gar nicht mehr unter den *Top Ten*). Der zuletzt gesellschaftspolitisch stark diskutierte Begriff *digital* (digitale Agenden, Digitalisierung etc.) taucht 2015 erstmals in den *Top Ten* auf.

2000	2001	2002	2003	2004	2005	2006	2007	2008
informati- on*	internet	internet	internet	elektro- nisch*	recht	elektro- nisch*	informati- on*	govern- ment
internet	elektroni- sche	elektro- nisch*	elektro- nisch*	govern- ment	voting	informati- on*	daten	informati- on
elektroni- sche	govern- ment	daten	govern- ment	voting	elektro- nisch*	recht	govern- ment	recht
wien	recht	govern- ment	recht	internet	online	govern- ment	voting	law
recht	online	domain	daten	recht	bereich	internet	internet	daten
computer	xml	verwal- tung	wien	verwal- tung	govern- ment	online	online	komplexi- tät
signaturen	informati- on	tammelo	österreich	bereich	frage	daten	recht	eu
demokra- tie	daten	recht	voting	informati- on	daten	form	services	legal
verwal- tung	software	xml	verwal- tung	wien	internet	google	web	online
ecash	commerce	commerce	schutz	daten	stellt	wissen	software	elektroni- sche

2009	2010	2011	2012	2013	2014	2015	2016
rechtsinfor- matik	daten	daten	law	information	data	law	legal
recht	law	legal	legal	daten	legal	legal	information
legal	legal	data	data	data	information	data	data
information	government	recht	internet	legal	law	information	law
fiedler	recht	law	software	law	daten	daten	daten
web	voting	government	recht	software	internet	digital	recht
informatik	information	information	information	recht	open	recht	online
law	online	internet	online	online	protection	internet	internet
wien	elektroni- sche	web	daten	form	recht	online	eu
juristische	open	wien	business	government	transparenz	knowledge	security

Tabelle 2: Die jeweils zehn häufigsten inhaltstragenden Begriffe des IRIS¹⁰

Die aufgrund der Analyse der Themengruppen festgestellten Trends zeigen sich auch bei der Vokabularanalyse. So findet sich der Begriff *government* in den ersten Jahren nur in den Jahren 2000, 2009 und 2012 nicht unter den zehn häufigsten Begriffen, während er zuletzt schon das dritte Jahr hintereinander (2014 bis 2016) nicht unter den zehn häufigsten Begriffen ist. Gleichzeitig war auch der Begriff *Verwaltung* nur 2000 sowie 2002 bis 2004 jeweils unter den *Top Ten*. Begriffe, die nur *einmal* unter den *Top Ten* zu finden sind, zeigen Besonderheiten dieser Bände auf, z.B. bei dem Band aus dem Jahr 2009 den Umstand, dass er HERBERT FIEDLER gewidmet ist, oder bei dem IRIS-Band 2014, der als Generalthema *Transparenz* hat.

Tabelle 3 zeigt, welche Begriffe am häufigsten in den *Top Ten* vertreten sind, wobei nur die Begriffe berücksichtigt sind, die wenigstens zweimal in den *Top Ten* vorkommen.

¹⁰ * = Trunkierungsoperator für zusammengeführte Vollformen (elektronisch {e, en}).

Begriff	Anzahl Jahre	Begriff	Anzahl Jahre	Begriff	Anzahl Jahre
recht	17	law	9	web	3
daten	15	legal	9	bereich	2
information*	14	data	6	commerce	2
internet	13	voting	5	eu	2
government	11	wien	5	form	2
online	10	software	4	open	2
elektronisch*	9	verwaltung	4	xml	2

Tabelle 3: Inhaltstragende Begriffe, die mindestens zweimal unter den zehn häufigsten (*Top Ten*) waren

Seit einigen Jahren hat das IRIS regelmäßig ein prägnantes Konzept bzw. ein Generalthema als Tagungsmotto, anders als in den Anfangsjahren, als längere Titel üblich waren. In den Jahren 2013 bis 2016 waren dies die Themen *Abstraktion und Applikation*, *Transparenz*, *Kooperation* und *Netzwerke* mit den jeweiligen englischen Entsprechungen. Wir haben durch einen Vergleich der Häufigkeiten der Begriffe im Volltext des jeweiligen Jahres untersucht, wie stark das jeweilige Generalthema auch im Inhalt des Tagungsbandes aufscheint, wobei wir uns auf die deutschsprachigen Begriffe konzentrieren. Dazu haben wir die häufigsten 300 inhaltstragenden Wörter in den jeweiligen Tagungsbänden untersucht. Der Effekt ist deutlich: So tauchen die Begriffe *Abstraktion* und *Applikation* (Generalthema 2013) weder in den Vorjahren 2010 bis 2012 noch in den Folgejahren 2014 bis 2016 unter den häufigsten 300 sinntragenden Begriffen auf, belegen aber 2013 die Plätze 56 (*Abstraktion*) und 274 (*Applikation*). Das Generalthema *Transparenz* aus dem Jahr 2014, das in diesem Jahr unter den *Top Ten* der häufigsten Begriffe ist, kommt in den beiden Folgejahren nicht unter die 300 häufigsten Begriffe und belegt in den Vorjahren 2010 bis 2013 lediglich im Jahr 2011 Rang 278. Ähnlich ist es bei dem Begriff *Kooperation*, der Rang 15 im Jahr dieses Generalthemas belegt und sonst in den Jahren 2010 bis 2016 nicht unter den 300 häufigsten Begriffen ist. Auch der Begriff *Netzwerke* bzw. *Netzwerke* (Generalthema im Jahr 2016) kommt 2016 unter den 300 häufigsten Begriffen vor (*Netzwerke* auf Rang 204 und *Netzwerk* auf Rang 277), während die beiden Begriffe in den Vorjahren 2010 bis 2015 nicht unter den 300 häufigsten sinntragenden Begriffen waren. Die Autoren greifen das jeweilige Tagungsmotto also klar erkennbar auf.

In Tabelle 4 sind die jeweils zehn auffälligsten Begriffe (*most distinctive words*) eines einzelnen Jahres im Verhältnis zum Gesamtkorpus aller IRIS-Bände aufgeführt. Die Jahre 2007 und 2008 fehlen, da aufgrund einer fehlerhaften Codierung und damit Fehlern im automatischen Datenexport die Daten für diese Auswertung nicht verwertbar sind. Anders als bei den häufigsten Begriffen, die eine Makroperspektive darstellen und eher die übergreifenden Themen herausarbeiten, zeigen diese Begriffe Besonderheiten der einzelnen Jahre auf: Beispielsweise war nur im Jahr 2000 von der *agora* die Rede, 2002 wurden *webbugs* thematisiert, ab 2011 taucht mehrfach *netzneutralität* als Thema auf und 2014 wurden *bitcoins* diskutiert.

2000	ecash, agora, tammelo's, iusmodell, münzen, spambots, zertifizierungsdiensteanbieter, sachverhaltsmodell, gegenakteur, schmutzer
2001	pte, vorabantrag, enderledigung, validator, informationsbegriff, watermarking, wasserzeichens, wasserzeichen, eugvü, raumzeitliche
2002	webbugs, webbug, schreiner, denic, nic.at, vergabestellen, starl, wbt, indexsuche, lernstext
2003	publikationsverzeichnis, öim, grundbuchs, argumentationsanalyse, cctld, webradio, rückruf, webradios, adoamt, unsend
2004	proefflokaal, kmdl, wien.at, bildung.at, stwt, auslandschweizer, parteiwechsel, tkv, raad, bildungspool
2005	enum, eumvvo, telfi, tabus, mahnverfahrens, gedankenexperimente, vsa, wfff, tabu, wissenslandkarte
2006	pitamic, jurex, vanity, reorganization, gpla, ersb, nummerninhaber, bsc, cecic, grundnorm
2009	smartvote, settlement, securities, trades, machinima, lexfind, vat, h.f, legiversum, xdoema
2010	örl, alrl, biometric, signatorin, strafwürdiger, productization, 2030, compliance, virtuellife, vorsitz
2011	dlp, ebi, netzneutralität, dyonipos, aphasie, scraping, gtelg, grids, metering, fcra
2012	noosphäre, facebook, zitiervorschlag, deceased, plagiarism, escrow, spyware, jusletter, pflegedokumentation, versorgungsleistungen
2013	usedsoft, aal, 9001, hörsaal, app, argumentationstheoriebasierte, psi, dpas, erschöpfungsgrundsatz, wahlbezirk
2014	pias, acta, bitcoins, osint, graduated, nichtrealistisches, zitiervorschlag, wdk, transparency, unbewegtes

2015	oss, cloud, parodie, qentis, fm, icj, contract, monkey, dpms, aktionärsforum
2016	iot, cyberlaw, bea, esma, cahve, hörsaal, policing, haapio, netzneutralität, kausalhaftung

Tabelle 4: Auffälligste Begriffe in den jeweiligen IRIS-Bänden

7. Fazit und Ausblick

20 Jahre IRIS: Im Ergebnis beobachten wir ein dialektisches Zusammenspiel von auch langfristig stabilen Kernthemen wie E-Government (wenn auch mit leichtem Rückgang), Rechtsinformation und IT-Recht einerseits und kurzfristigeren Trends, die nur einige Jahre das IRIS prägen, andererseits (z.B. elektronisches Publizieren in den Anfangsjahren oder Netzneutralität in den letzten Jahren). Die Visualisierung häufigster Begriffe für den ältesten (2000) bzw. den neuesten Band (2016) als Wortwolke betont die Änderungen (Abbildung 5): Weder *elektronisch* noch überhaupt der *Computer* scheint in der rechten Wolke auf, die dafür stärker von (englischen) Kernbegriffen geprägt ist.



Abbildung 5: Wortwolken häufigster Begriffe des IRIS 2000 (links) bzw. 2016 (rechts)

Die breite thematische Vielfalt des IRIS zeigt die für das Gesamtcorpus der vorliegenden Bände berechnete Wortwolke der 500 häufigsten Begriffe (Abbildung 6).

Die hier präsentierten Daten stellen nur einen kleinen Ausschnitt aus dem gewonnenen Material dar. Insofern bleibt die Frage, wie Ergebnisse solcher Analysen aggregiert und zugänglich gemacht werden können: Kann es gelingen, dies in verdichteter (und standardisierter) statischer Form als Aufsatz oder Bericht zusammenzufassen? Oder sollte man nicht davon ausgehen, dass das Ergebnis einer solchen Analyse ein Informationssystem sein sollte, das von unterschiedlichen Nutzern mit unterschiedlichen Interessen abgefragt werden kann?¹¹

Die Entscheidung für einfache begriffsbezogene Metriken – letztlich Häufigkeitsstatistiken des verwendeten Vokabulars – soll nicht darüber hinwegtäuschen, dass weitere Analyseebenen, die komplexere linguistische Strukturen berücksichtigen, ebenfalls sinnvoll wären: Hier ist z.B. an die Untersuchung von Argumentationsmustern anhand typischer Satzbaumuster zu denken. Vorliegend könnte man untersuchen, ob sich Argumentationsketten in der Rechtsinformatik im Zeitverlauf verändert haben. Eine ähnliche Betrachtung könnte anhand der Analyse der Aufbaustrukturen der wissenschaftlichen Aufsätze erfolgen. Gerade in den Natur- und Lebenswissenschaften haben sich mittlerweile standardisierte Aufbaustrukturen etabliert, die typischerweise von wissenschaftlichen Beiträgen erwartet werden, und deren Einhaltung als Qualitätsmerkmal gilt. Zu fragen wäre, ob eine vergleichbare Standardisierung oder Vereinheitlichung auch in der Rechtsinformatik zu beobachten ist. Auf einer interpretativen Ebene wäre ergänzend zu diskutieren, ob solche Aufbauschemata überhaupt sinnvoll anwendbar sind auf ein so heterogenes und polymethodisch geprägtes Fachgebiet wie die Rechtsinformatik. Zu fragen ist auch, ob es bibliometrische Maße gibt, die die inhaltliche Veränderung eines Themenfelds adäquat beschreiben, und wie solche Maße aussehen könnten.

¹¹ Vgl. SIPPL/BURGHARDT/WOLFF/MIELKE 2016 (Informationssystem zu österreichischen Parlamentsreden).

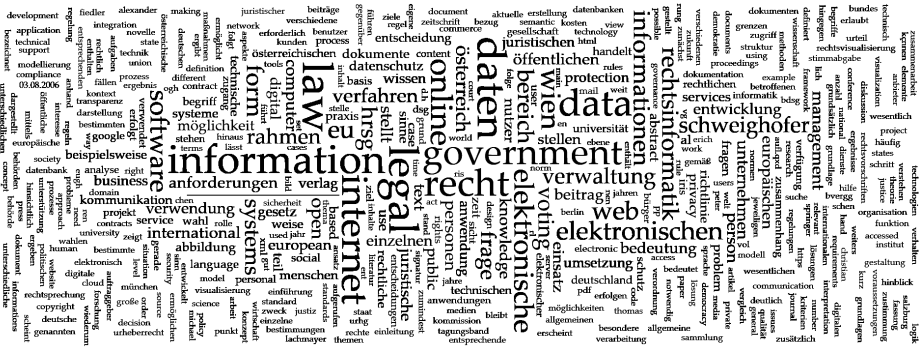


Abbildung 6: Wortwolke der 500 häufigsten Begriffe im Gesamtkorpus aller IRIS-Bände

8. Danksagung

Dieser Beitrag wäre nicht möglich gewesen ohne die Unterstützung durch Herausgeber und Verleger des IRIS. Wir sind daher Franz Kummer und Erich Schweighofer zu großem Dank verpflichtet, die keine Mühe gescheut haben, die elektronischen Artikel der IRIS-Tagungsbände verfügbar zu machen. Ein besonderer Dank gilt Anne Helms (Weblaw AG, Bern) für die technische Unterstützung bei der Bereitstellung der Texte.

9. Literatur

BALL, RAFAEL, *Bibliometrie, Einfach – verständlich – nahvollziehbar*, de Gruyter, Berlin/Boston 2014.

CRETCHLEY, JULIA/ROONEY, DAVID/GALLOIS, CINDY, *Mapping a 40-Year History With Leximancer: Themes and Concepts in the Journal of Cross-Cultural Psychology*, *Journal of Cross-Cultural Psychology* 2010, 41 (3), S. 318–328. DOI:10.1177/0022022110366105.

GARFIELD, EUGENE, *Citation Indexes for Science, A New Dimension in Documentation through Association of Ideas*, *Science* 1955, 122, Nr. 3159, S. 108–111.

MIELKE, BETTINA/WOLFF, CHRISTIAN, *Österreichisch-deutsche Rechtssprache kontrastiv. Eine corpuslinguistische Analyse*. In: Schweighofer, Erich/Kummer, Franz/Hötzendorfer, Walter (Hrsg.), *Abstraktion und Applikation, Tagungsband des 16. Internationalen Rechtsinformatik Symposions IRIS 2013*, books@ocg.at, Wien 2013, S. 377–184.

MIELKE, BETTINA/WOLFF, CHRISTIAN, *Justiz und Digitale Öffentlichkeit: Aufbau und Analyse eines Twittercorpus zum Thema Justiz*. In: Schweighofer, Erich/Kummer, Franz/Hötzendorfer, Walter (Hrsg.), *Kooperation, Tagungsband des 18. Internationalen Rechtsinformatik Symposions IRIS 2015*, books@ocg.at, Wien 2015, S. 281–290.

MIELKE, BETTINA/WOLFF, CHRISTIAN, *Österreichische und Deutsche Gerichtsentscheidungen im Sprachvergleich*. In: Schweighofer, Erich/Kummer, Franz/Hötzendorfer, Walter/Borges, Georg (Hrsg.), *Netzwerke, Tagungsband des 19. Internationalen Rechtsinformatik Symposions IRIS 2016*, books@ocg.at, Wien 2016, S. 129–138.

SINCLAIR, STÉFAN/ROCKWELL, GEOFFREY/THE VOYANT TOOLS TEAM, *Voyant Tools* (web application), 2012, <https://voyant-tools.org/>.

SIPPL, COLIN/BURGHARDT, MANUEL/MIELKE, BETTINA/WOLFF, CHRISTIAN, *Korpusbasierte Analyse österreichischer Parlamentsreden*. In: Schweighofer, Erich/Kummer, Franz/Hötzendorfer, Walter/Borges, Georg (Hrsg.), *Netzwerke, Tagungsband des 19. Internationalen Rechtsinformatik Symposions IRIS 2016*, books@ocg.at, Wien 2016, S. 139–148.

RAVIKUMAR, S./AGRAHARI, ASHUTOSH/SINGH, S. N., *Mapping the intellectual structure of scientometrics: a co-word analysis of the journal Scientometrics (2005–2010)*, *Scientometrics* 2015, 102 (1), S. 929–955. DOI:10.1007/s11192-014-1402-8.

ROONEY, DAVID/MCKENNA, BERNARD/BARKER, JAMES R., *History of Ideas in Management Communication Quarterly*, *Management Communication Quarterly* 2011, 20 (10), S. 1–29. DOI:10.1177/0893318911405623.

SIDOROVA, ANNA/EVANGELOPOULOS, NICHOLAS/VALACICH, JOSEPH S./RAMAKRISHNAN, THIAGARAJAN, *Uncovering the intellectual core of the information systems discipline*, *MIS Quarterly* 2008, 32 (3), S. 467–482.