

Atıf Klasiklerinin Etkisinin ve İlgililik Sıralamalarının Pennant Diyagramları ile Analizi

Müge Akbulut

4. Uluslararası Yönetim Bilişim Sistemleri Konferansı (YBS'17)
Doktora Konsorsiyumu. 17-20 Ekim 2017



Plan

- Sorun
- Hipotezler
- Yöntem
- Pennant diyagramları ve ilgililik sıralamaları
- Bulgular ve sonuç

Sorun

- Klasiklerin etkisini izleme
- İlgililik sıralamaları
- Alanlar arası ilişkilerin görselleştirilmesi
- Maron ve Kuhns, 1960

Maron ve Kuhns, 1960

- Olasılıksal bilgi erişim
- Sorgu genişlemesi
- Arama stratejileri
- Ampirik verilere dayalı sonuçlar

On Relevance, Probabilistic Indexing and Information Retrieval*

M. E. MARON

The RAND Corporation, Santa Monica, California

AND

J. L. KUHN

Ramo-Wouldridge, Canoga Park, California

Abstract. This paper reports on a novel technique for literature indexing and searching in a mechanized library system. The notion of relevance is taken as the key concept in the theory of information retrieval and a comparative concept of relevance is explicated in terms of the theory of probability. The resulting technique called "Probabilistic Indexing," allows a computing machine, given a request for information, to make a statistical inference and derive a number (called the "relevance number") for each document, which is a measure of the probability that the document will satisfy the given request. The result of a search is an ordered list of those documents which satisfy the request ranked according to their probable relevance.

The paper goes on to show that whereas in a conventional library system the cross-referencing ("see" and "see also") is based solely on the "semantical closeness" between index terms, statistical measures of closeness between index terms can be defined and computed. Thus, given an arbitrary request consisting of one (or many) index term(s), a machine can elaborate on it to increase the probability of selecting relevant documents that would not otherwise have been selected.

Finally, the paper suggests an interpretation of the whole library problem as one where the request is considered as a clue on the basis of which the library system makes a concatenated statistical inference in order to provide as an output an ordered list of those documents which most probably satisfy the information needs of the user.

Hipotezler

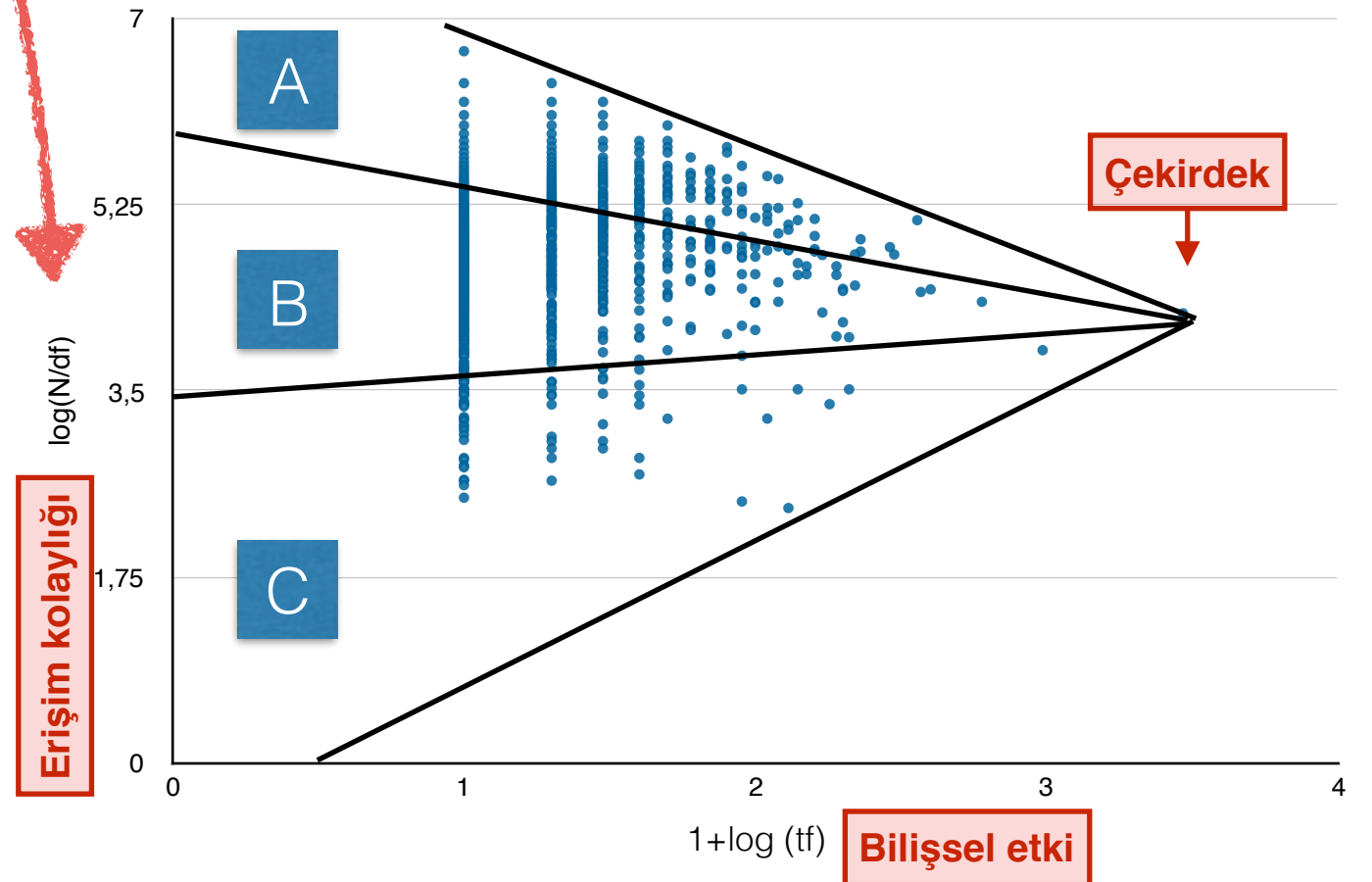
- Geleneksel atıf analizi ile gözlenemeyen disiplinlerarası ilişkiler Pennant diyagramları yöntemi ile ortaya çıkarılabilir
- Ortak atıf sayılarına göre oluşturulan ilgililik sıralamaları kaynakça benzerliğine göre oluşturulan ilgililik sıralamasından daha başarılıdır

Pennant Diyagramları

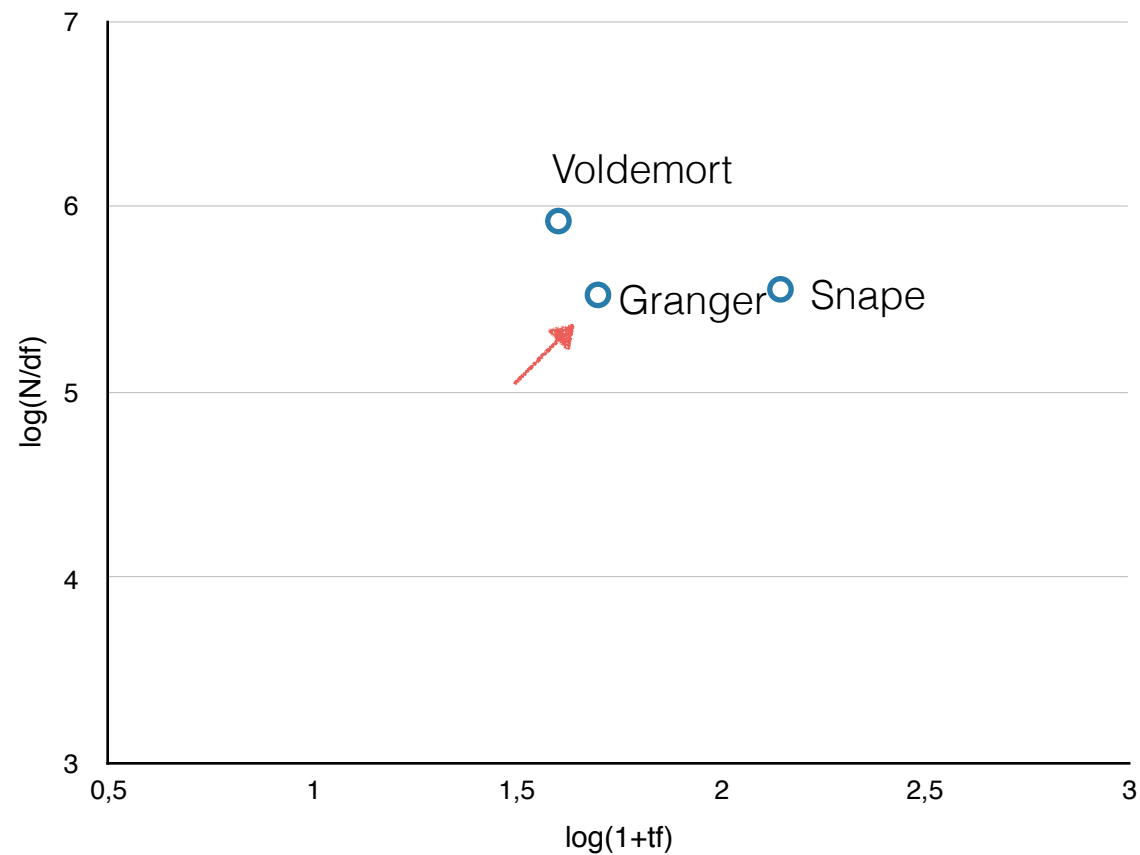
- Ortak Atıf (co-citation)
- Ters Belge Sıklığı (Sparck Jones, 1973)
- “İlgililik = bilişsel etki/ işleme çabası” (Sperber ve Wilson, 1995)
- Pennant diyagramları; bibliyometrideki ortak atıf, bilgi erişimdeki $tf*idf$ ve ilgililik kuramını birleştirmektedir.

Pennant Diyagramları

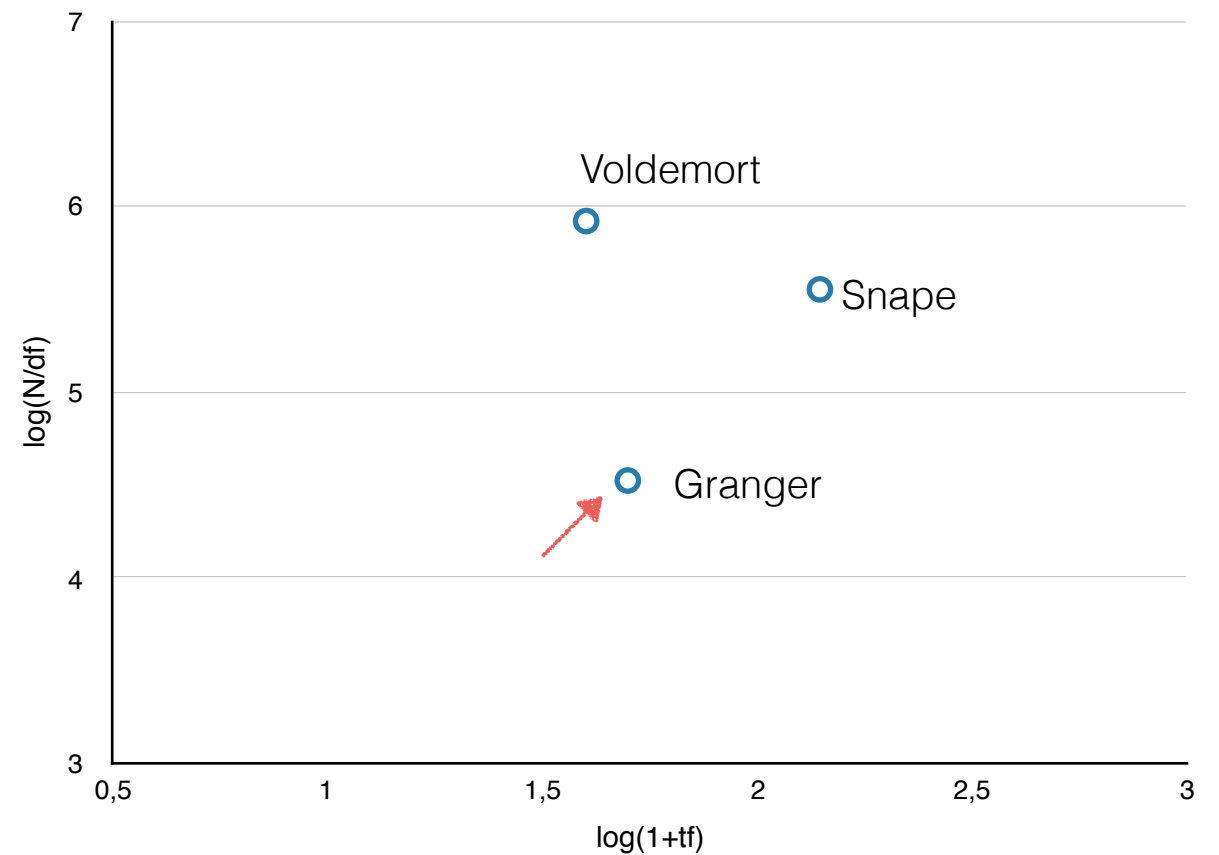
	tf	df	1+log(tf)	log(N/df)	Ağırlık (1+log(tf))*(og(N/df))
Maron, 1960	294	294	3,468347	4,230622	14,67326886
Cooper, 1978	36	50	2,556302	5	12,7815125
Robertson, 1977	45	112	2,653212	4,649751	12,33678014
Robertson, 1976	60	237	2,778151	4,324221	12,01334181
Croft, 1979	97	654	2,986771	3,883392	11,59880622
Robertson, 1994	12	16	2,079181	5,494850	11,42478911
Salton, 1983	12	16	2,079181	5,494850	11,42478911
Bookstein, 1977	16	38	2,204119	5,119186	11,28330106
Winkler, 1968	11	15	2,041392	5,522878	11,27436427
Cooper, 1978	9	12	1,954242	5,619788	10,98243009



Pennant Diyagramları

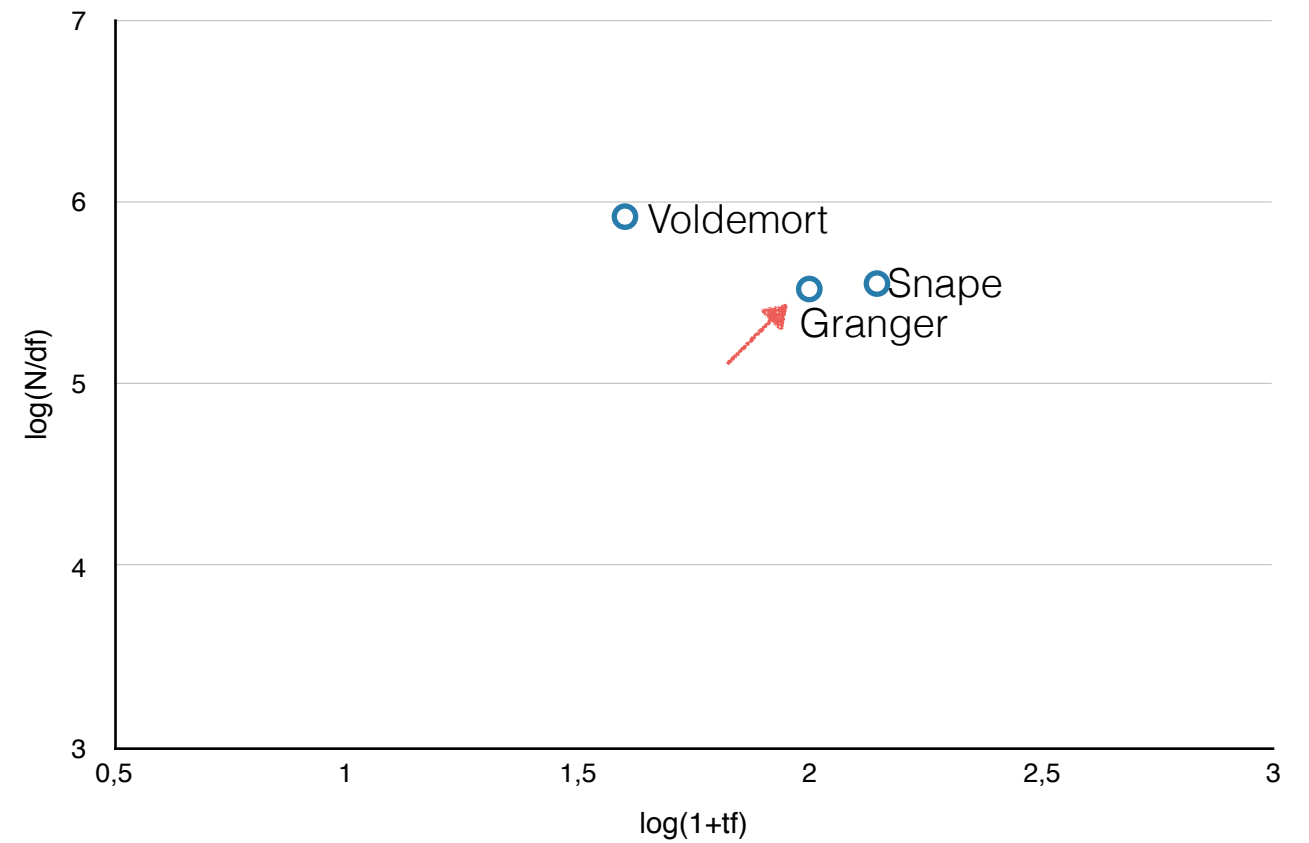
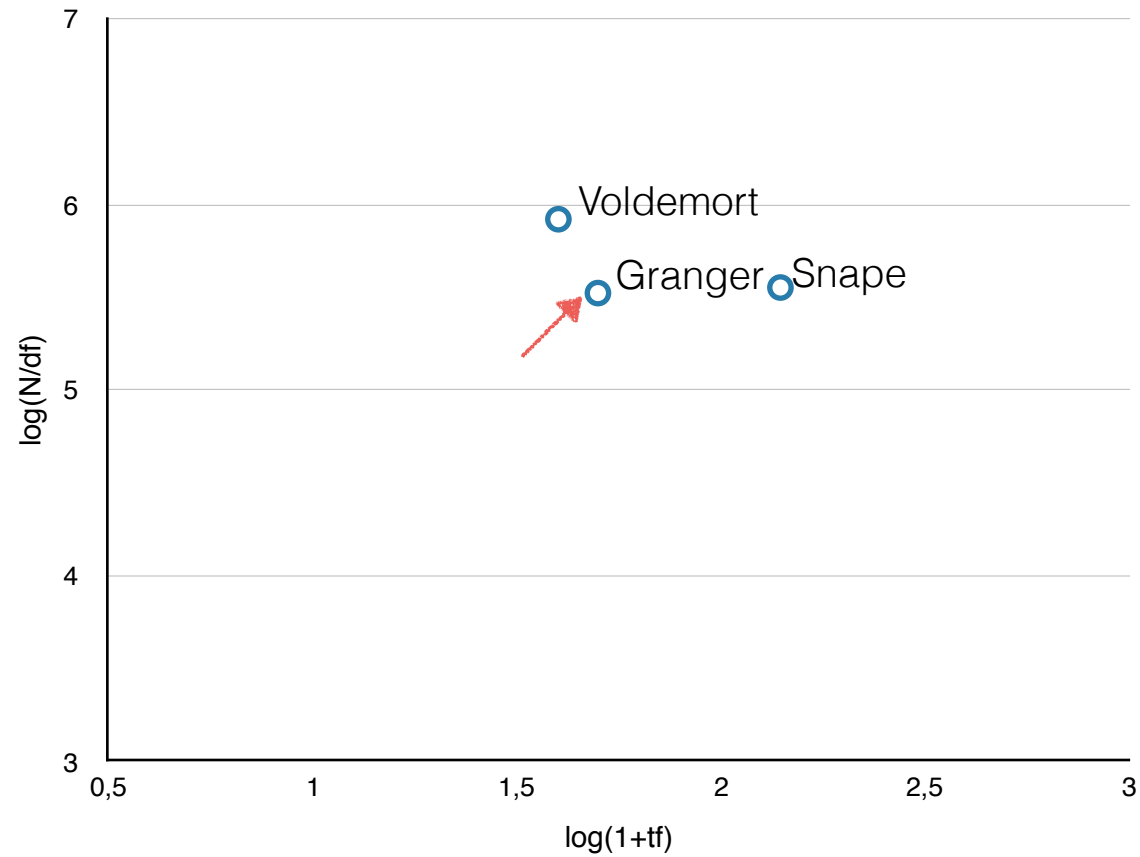


	df	tf	weight	$\log(tf)$	$\log(idf)$
	df	tf	$tf \cdot idf$	$\log(1+tf)$	$\log(5mil/df)$
Snape	14	14	11,91710983	2,146128036	5,552841969
Granger	15	5	9,383205326	1,698970004	5,522878745
Voldemort	6	4	9,485506842	1,602059991	5,920818754



	df	tf	weight	$\log(tf)$	$\log(idf)$
	df	tf	$tf \cdot idf$	$\log(1+tf)$	$\log(5mil/df)$
Snape	14	14	11,91710983	2,146128036	5,552841969
Granger	150	5	7,684235321	1,698970004	4,522878745
Voldemort	6	4	9,485506842	1,602059991	5,920818754

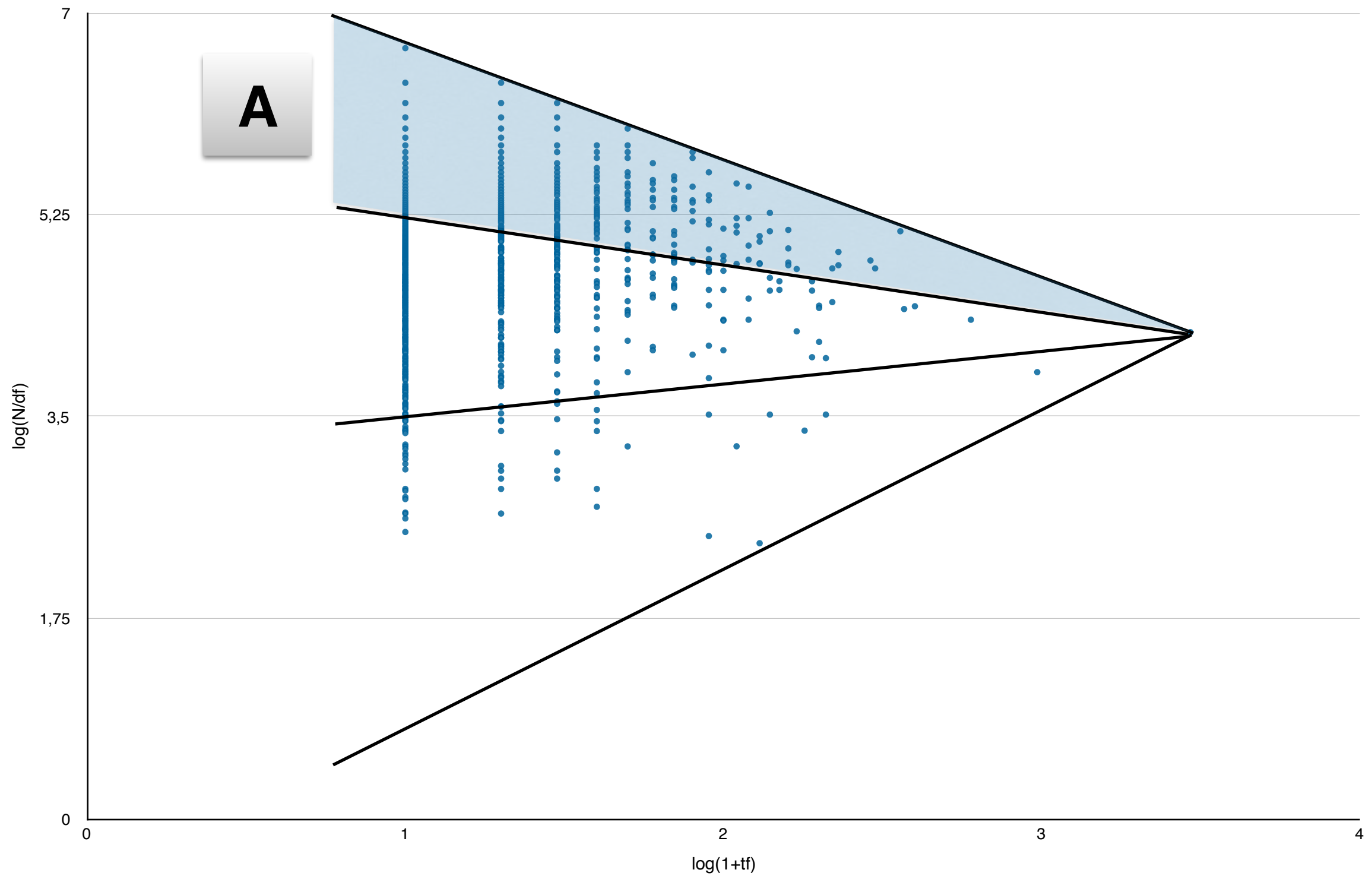
Pennant Diyagramları



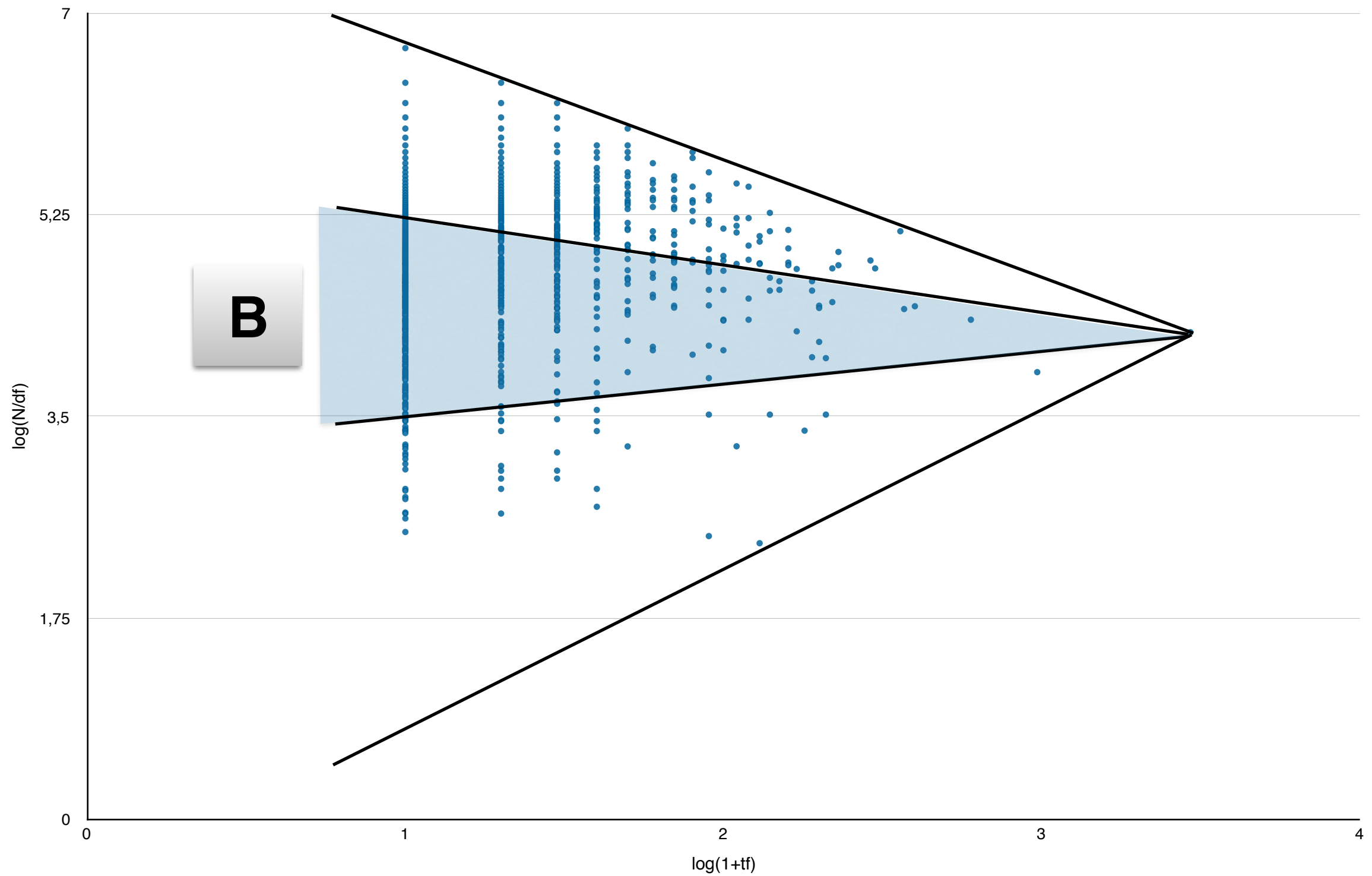
			weight	$\log(tf)$	$\log(idf)$
	df	tf	tf*idf	$\log(1+tf)$	$\log(5mil/df)$
Snape	14	14	11,91710983	2,146128036	5,552841969
Granger	15	5	9,383205326	1,698970004	5,522878745
Voldemort	6	4	9,485506842	1,602059991	5,920818754

			weight	$\log(tf)$	$\log(idf)$
	df	tf	tf*idf	$\log(1+tf)$	$\log(5mil/df)$
Snape	14	14	11,91710983	2,146128036	5,552841969
Granger	15	10	11,04575749	2	5,522878745
Voldemort	6	4	9,485506842	1,602059991	5,920818754

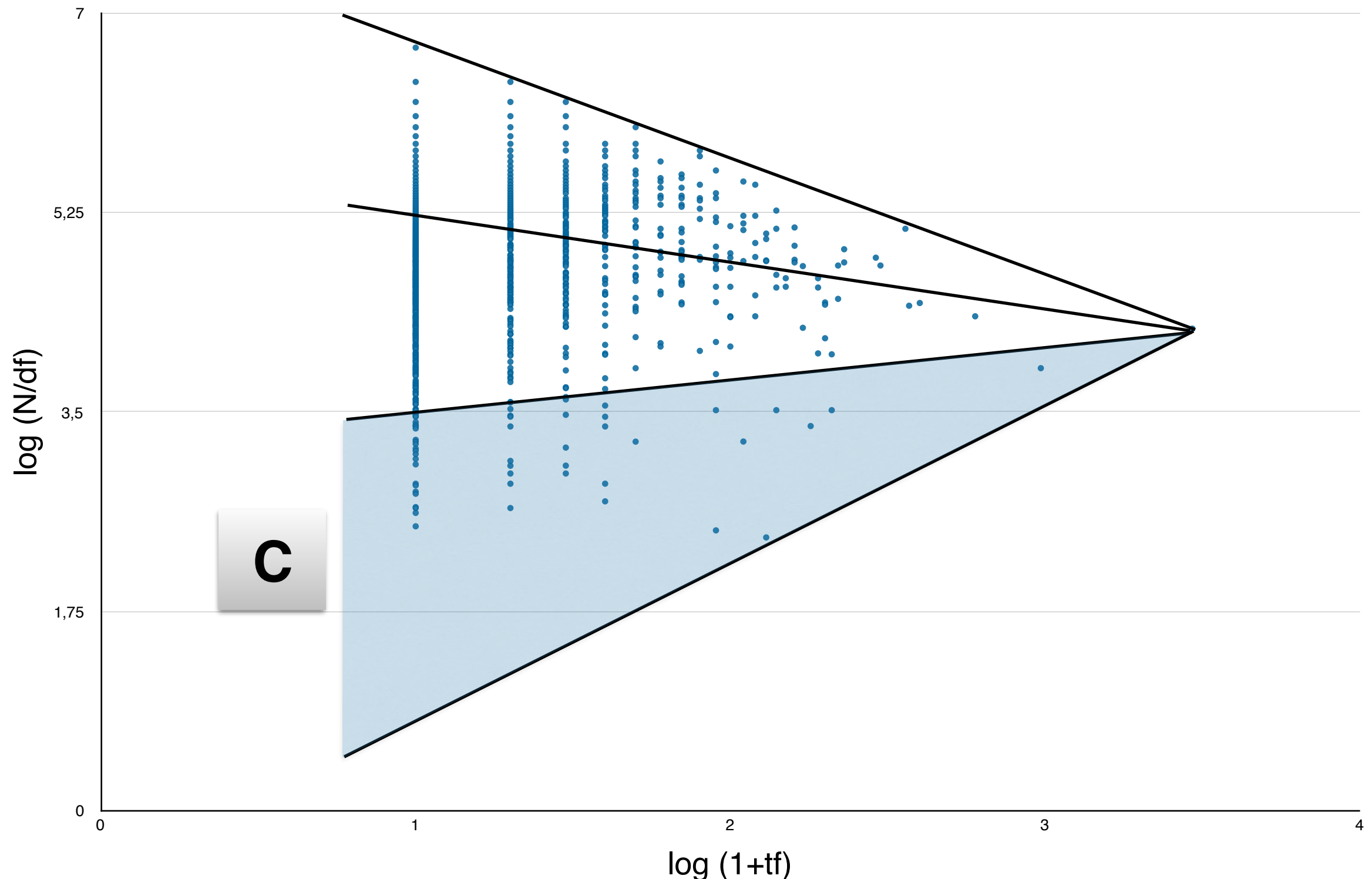
Sektörler



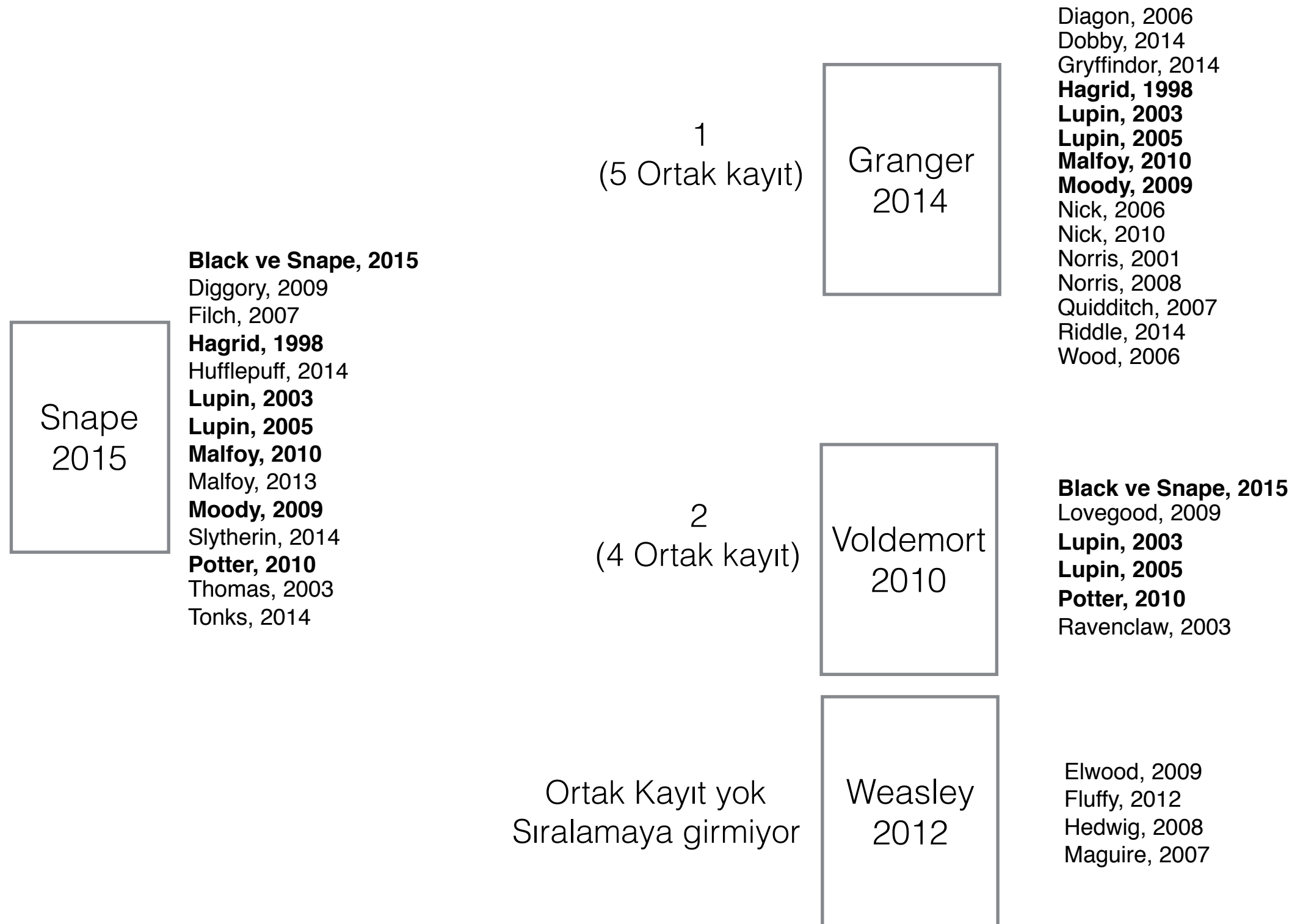
Sektörler



Sektörler



WoS İlgili Kayıtlar Özelliği



WoS İlgili Kayıtlar Özelliği

[Web of Science](#) [InCites](#) [Journal Citation Reports](#) [Essential Science Indicators](#) [EndNote](#) [Publons](#) [Sign In](#) [Help](#) [English](#)

Web of Science 

[Search](#) [Search Results](#) [My Tools](#) [Search History](#) [Marked List](#)

[Full Text from Publisher](#)   [Save to EndNote online](#) [Add to Marked List](#) 2 of 2

ON RELEVANCE, PROBABILISTIC INDEXING AND INFORMATION RETRIEVAL

By: [MARON, ME](#) (MARON, ME); [KUHN, JL](#) (KUHN, JL)

JOURNAL OF THE ACM
Volume: 7 Issue: 3 Pages: 216-244
DOI: 10.1145/321033.321035
Published: 1960
[View Journal Impact](#)

Publisher
ASSOC COMPUTING MACHINERY, 1515 BROADWAY, NEW YORK, NY 10036

Categories / Classification
Research Areas: Computer Science
Web of Science Categories: Computer Science, Hardware & Architecture; Computer Science, Information Systems; Computer Science, Software Engineering; Computer Science, Theory & Methods

Document Information
Document Type: Article
Language: English
Accession Number: WOS:A1960WV56100002
ISSN: 0004-5411

Citation Network

320 Times Cited
[2 Cited References](#)
[View Related Records](#)
[Create Citation Alert](#)
(data from Web of Science Core Collection)

All Times Cited Counts

327 In All Databases
320 In Web of Science Core Collection
4 In BIOSIS Citation Index
7 In Chinese Science Citation Database
0 In Data Citation Index
0 In Russian Science Citation Index
1 In SciELO Citation Index

Usage Count

Last 180 Days: 1
Since 2013: 24

WoS İlgili Kayıtlar Özelliği

[Web of Science](#) [InCites](#) [Journal Citation Reports](#) [Essential Science Indicators](#) [EndNote](#) [Publons](#) [Sign In](#) [Help](#) [English](#)

Web of Science 

[Search](#) [Search Results](#) [My Tools](#) [Search History](#) [Marked List](#)

Related Records: 10,627
(from Web of Science Core Collection)
For: ON RELEVANCE, PROBABILISTIC INDEXING AND INFORMATION RETRIEVAL ...[More](#)

Refine Results
 
Filter results by:
☐ Highly Cited in Field (29) 
☐ Hot Papers in Field (1) 
[Refine](#)
Publication Years
☐ 2015 (543)
☐ 2016 (512)
☐ 2014 (507)
☐ 2012 (501)
☐ 2013 (494)
[more options / values...](#)

Sort by: Relevance
Page 1 of 1,063
☐ Select Page   5K [Add to Marked List](#) Citation Report feature not available. [\[?\]](#)
[Analyze Results](#)

☐ 1. **Relationship of non-specific commensalism in the colonization of the deep layers of sediment**
By: Prevedelli, D; Simionini, R; Ansaloni, I
JOURNAL OF THE MARINE BIOLOGICAL ASSOCIATION OF THE UNITED KINGDOM Volume: 81 Issue: 6
Pages: 897-901 Published: DEC 2001
[View Abstract](#)
Times Cited: 9
(from Web of Science Core Collection)
Cited References: 23
Shared References: 2
Usage Count 

☐ 2. **A novel mobile device user interface with integrated social networking services**
By: Cui, Yanqing; Honkala, Mikko
INTERNATIONAL JOURNAL OF HUMAN-COMPUTER STUDIES Volume: 71 Issue: 9 Pages: 919-932
Published: SEP 2013
[Full Text from Publisher](#) [View Abstract](#)
Times Cited: 7
(from Web of Science Core Collection)
Cited References: 53
Shared References: 1
Usage Count 

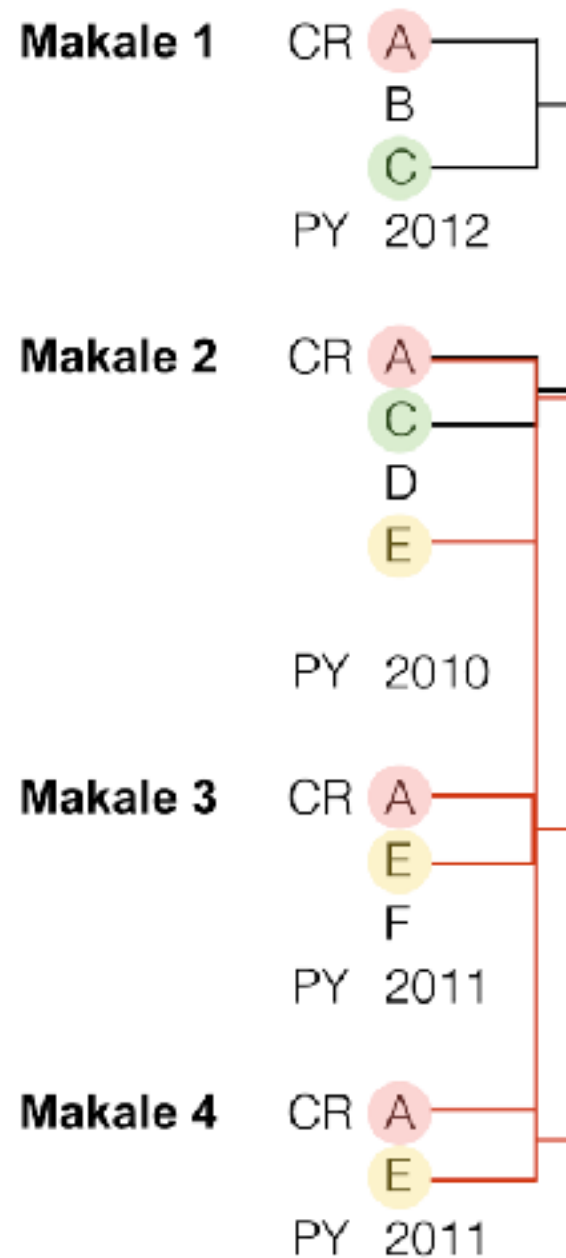
☐ 3. **Significant Prognostic Features and Patterns of Somatic TP53 Mutations in Human Cancers**
By: Zhang, Wensheng; Edwards, Andrea; Flemington, Erik K.; et al.
CANCER INFORMATICS Volume: 16 Article Number: 1176935117691267 Published: FEB 20 2017
[Full Text from Publisher](#) [View Abstract](#)
Times Cited: 0
(from Web of Science Core Collection)
Cited References: 56
Shared References: 1

Yöntem

Veri Seti

- Çekirdek makaleye atıf yapan çalışmaların kaynakçaları
- 294 çalışma, 9607 referans (tekil sayı ise 4176)

Ortak Atıf Hesaplama



Referans	2010	2011	2012
A	1	2	1
B	0	0	1
C	1	0	1
D	1	0	0
E	0	2	0
F	0	1	0

Standardizasyon ve Birleştirme

WoS CR formatı

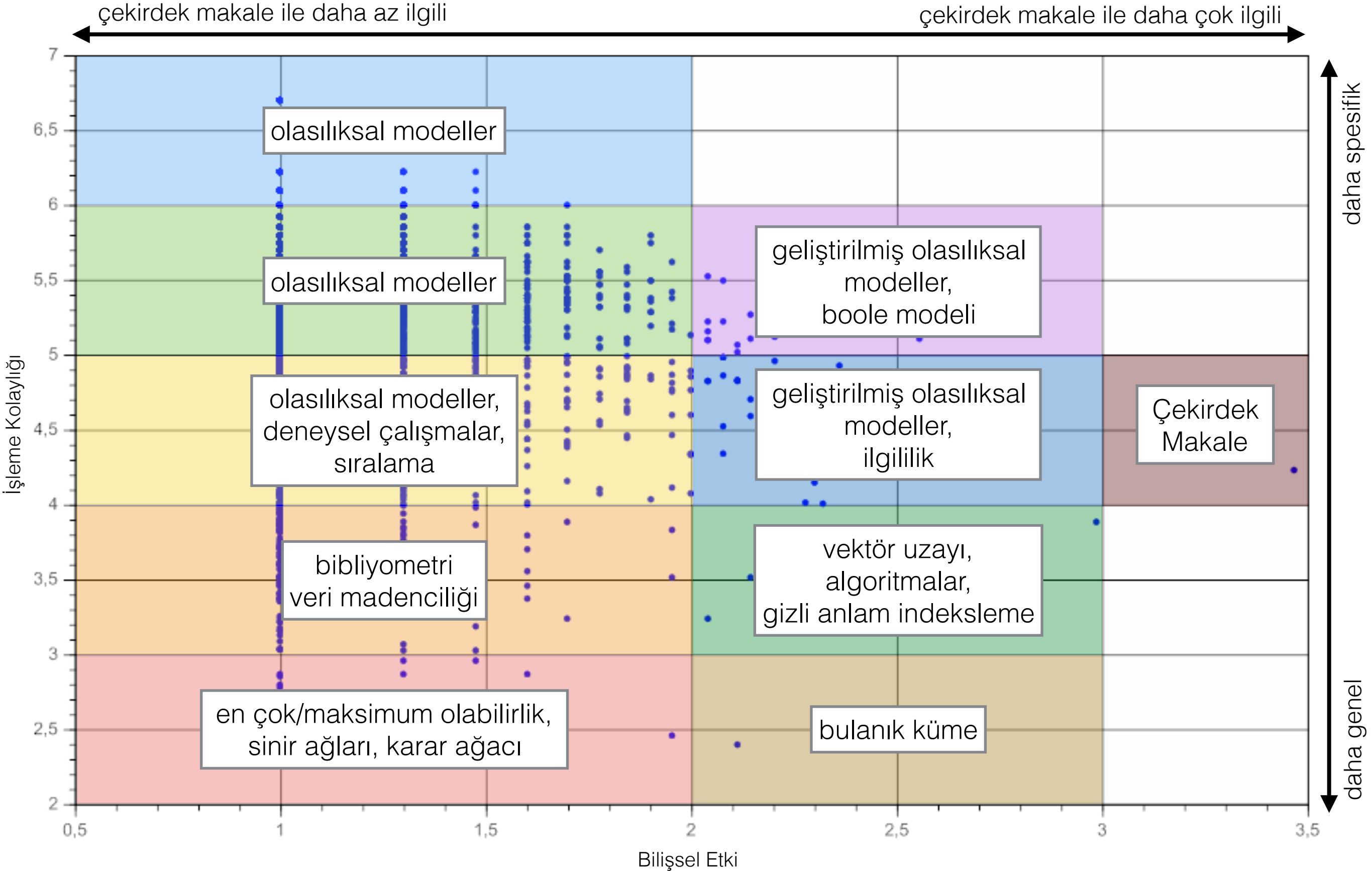
	A	B	C	D	E	F	G	H	I	J	K
1	TF_MACR_CR	TF_1960	TF_1961	TF_1962	TF_1963	TF_1964	TF_1965	TF_1966	TF_1967	TF_1968	TF_1969
2	KUIPERS JW, 1957, AM DOC, V8, P246, DOI 10.1002/asi.5090080403	1	0	0	0	0	0	0	0	0	0
3	MARON ME, 1961, J ACM, V8, P404, DOI 10.1145/321075.321084	0	0	2	0	0	3	1	0	0	2
4	SWANSON DR, 1960, SCIENCE, V132, P1099, DOI 10.1126/science.132.3434.1 099	1	0	1	1						0
5	Zhang C, 2010, IEEE NETWORK, V24, P13	0	0	0	0	0					0

dergi adı
kısaltma listesi

Atıf
Raporu

	A	B	C	D	E	F	G	H	I	J	K
1	IDF_Title	IDF_Authors	IDF_Source Title	IDF_Publication Year	IDF_Volume	IDF_Issue	IDF_Beginning Page	IDF_Ending Page	IDF_1960	IDF_1961	IDF_1962
2	A MINICARD SYSTEM FOR DOCUMENTARY INFORMATION	Kuipers, JW; Tyler, AW; Myers, WL	AMERICAN DOCUMENTATION	1957	8	4	246	268	2	0	2
3	AUTOMATIC INDEXING - EXPERIMENTAL INQUIRY	Maron, ME	JOURNAL OF THE ACM	1961	8	3	404	412	0	0	2
4	SEARCHING NATURAL LANGUAGE TEXT BY COMPUTER	Swanson, DR	SCIENCE	1960	132	3434	1099	1104	1	1	5
5	Privacy and Security for Online Social Networks: Challenges and Opportunities	Zhang, Chi; Sun, Jinyuan; Zhu, Xiaoyan; Fang, Yuguang	IEEE NETWORK	2010	24	4	13	18	0	0	0

Bulgular



Etkileşimli Grafik



İlgililik Sıralamaları

- Pennant yöntemi ile 4176, WoS ilgili kayıtlar yöntemi ile 9803
- İlk 50'şer kayıt
- Kapsama oranı %0

WoS RR - Pennant

Sıra No	WoS Kaynakça Benzerliğine Göre Oluşturulan İlgilik Sıralaması	Pennant Diyagramları Yöntemine Göre Oluşturulan İlgilik Sıralaması
1	Prevedelli, D., Simonini, R. ve Anselmi, I. (2001). Relationship of non-specific commensalism in the colonization of the deep layers of sediment. <i>Journal of the Marine Biological Association of the UK</i> , 81(06), 897-901.	Cooper, W. S. ve Maron, M. E. (1978). Foundations of probabilistic and utility-theoretic indexing. <i>Journal of the ACM (JACM)</i> , 25(1), 67-80.
2	Lucia, L., Lo, D., Jiang, L., Thung, F. ve Budi, A. (2014). Extended comprehensive study of association measures for fault localization. <i>Journal of Software: Evolution and Process</i> , 26(2), 172-219.	Robertson, S. E., Maron, M. E., & Cooper, W. S. (1982). Probability of relevance: A unification of two competing models for document retrieval. <i>Information Technology: Research and development</i> , 1(1), 1-21.
3	Tedeschini, R., Consonni, V., Xiang, H., Holliday, J., Buscema, M. ve Willert, P. (2012). Similarity coefficients for binary chemoinformatics data: Overview and extended comparison using simulated and real data sets. <i>Journal of Chemical Information and Modeling</i> , 52(11), 2884-2901.	Robertson, S. E. (1997). The probability ranking principle in IR. <i>Readings in information retrieval</i> , 281-286.
4	Ghosh, S., Carranza, E. J. M., Van Westen, C. J., Jetten, V. G. ve Bhattacharya, D. N. (2011). Selecting and weighting spatial predictors for empirical modeling of landslide susceptibility in the Darjeeling Himalayas (India). <i>Geomorphology</i> , 131(1), 35-56.	Robertson, S. E. ve Sparck Jones, K. (1976). Relevance weighting of search terms. <i>Journal of the American Society for Information Science</i> , 27(3), 129-146.
5	Tan, P. N., Kumar, V. ve Srivastava, J. (2004). Selecting the right objective measure for association analysis. <i>Information Systems</i> , 29(4), 293-313.	Croft, W. B., & Harper, D. J. (1979). Using probabilistic models of document retrieval without relevance information. <i>Journal of Documentation</i> , 35(4), 285-295.
6	Shoukri, M. M., Chaudhary, M. A. ve Mohamed, G. H. (2003). Evaluating normal approximation confidence intervals for measures of 2 x 2 association with applications to twin data. <i>Biometrical Journal</i> , 45(1), 20-33.	Robertson, S. E. (1994). Documentation note query-document symmetry and dual models. <i>Journal of Documentation</i> , 50(3), 233-238. Robertson, S. E. (1994). Documentation note query-document symmetry and dual models. <i>Journal of Documentation</i> , 50(3), 233-238.
7	Reboussin, B. A. ve Liang, K. Y. (1998). An estimating equations approach for the LISCOMP model. <i>Psychometrika</i> , 63(2), 165-182.	Salton, G., Fox, E. A. ve Wu, H. (1983). Extended Boolean information retrieval. <i>Communications of the ACM</i> , 26(11), 1022-1036.
8	Izmaylov, D. M. ve Obukhova, L. K. (1996). Geroprotector efficiency depends on viability of control population: Life span investigation in D - melanogaster. <i>Mechanisms of Ageing and Development</i> , 91(3), 155-164.	Bookstein, A. ve Kraft, D. (1977). Operations research applied to document indexing and retrieval decisions. <i>Journal of the ACM</i> , 24(3), 418-427.
9	Goodman, L. A. (1981). Association models and the bivariate normal for contingency tables with ordered categories. <i>Biometrika</i> , 68(2), 347-355.	Winkler, R. L. (1968). The consensus of subjective probability distributions. <i>Management Science</i> , 15(2), B-61-B-75.
10	Wilson E.B. (1964). Comparative Experiment + Observed Association .2. <i>Proceedings of The National Academy of Sciences of The United States of America</i> , 51(4), 539-542.	Cooper, W. S. (1978). Indexing documents by gedanken experimentation. <i>Journal of the American Society for Information Science</i> , 29(3), 107-119.
11	Evans, F. B. (1969). Psychological and objective factors in the prediction of brand choice ford versus chevrolet. <i>The Journal of Business</i> , 32(4), 340-369.	Bookstein, A. (1963). Outline of a general probabilistic retrieval model. <i>Journal of Documentation</i> , 39(2), 63-72.
12	Li, F., Chu, H., ve Nie, L. (2011). A two-stage estimation for screening studies using two diagnostic tests with binary disease status verified in test positives only. <i>Statistical Methods in Medical Research</i> , 24(6), 635-66.	Bookstein, A. ve Cooper, W. (1976). A general mathematical model for information retrieval systems. <i>The Library Quarterly</i> , 153-167.
13	Crusio, C. A., Cantzler, J., Roberts, Y. H., & Hooper, L. M. (2015). Psychometric Evaluation of Data From the Race-Related Events Scale. <i>Measurement and Evaluation in Counseling and Development</i> , 48(4), 285-296.	Luhn, H. P. (1981). The automatic derivation of information retrieval encodings from machine-readable texts. A. Kent (Yay. haz). <i>Information retrieval and machine translation</i> (Cilt 3, Bölüm 2, ss. 1021-1028) içinde. New York: Interscience Publication.
14	Alofs, K. M. ve Jackson, D. A. (2015). The vulnerability of species to range expansions by predators can be predicted using historical species associations and body size. <i>Proceedings of the Royal Society of London B: Biological Sciences</i> , 292(1812). doi: 10.1098/rspb.2015.1211	Fuhr, N. (1989). Models for retrieval with probabilistic indexing. <i>Information Processing and Management</i> , 22(1), 55-72.
15	Baker, S. G. (2015). RE: Combined Associations of Genetic and Environmental Risk Factors: Implications for Prevention of Breast Cancer. <i>Journal of the National Cancer Institute</i> , 107(6), cjr127.	Bookstein, A. (1978). On the perils of merging Boolean and weighted retrieval systems. <i>Journal of the American Society for Information Science</i> , 29(3), 156-158.
16	Yousefi, M. ve Carranza, E. J. M. (2015). Prediction-area (P-A) plot and C-A fractal analysis to classify and evaluate evidential maps for mineral prospectivity modeling. <i>Computers and Geosciences</i> , 79, 69-81.	Croft, W. B. ve Lafferty, J. (Yay. haz.). (2003). <i>Language modeling for information retrieval</i> . Boston: Kluwer Academic.
17	Negri, M. ve Sprumont, Y. (2015). Size invariant measures of association: Characterization and difficulties. <i>Mathematical Social Sciences</i> , 75, 115-122.	Bookstein, A. (1979). Relevance. <i>Journal of the American Society for Information Science</i> , 30(5), 269-273.

Sonuç

- Literatür izleme
- Başarılı ilgililik sıralamaları
- İlişkiler ve zamana göre değişim

Atıf Klasiklerinin Etkisinin ve İlgililik Sıralamalarının Pennant Diyagramları ile Analizi

Müge Akbulut

4. Uluslararası Yönetim Bilişim Sistemleri Konferansı (YBS'17)
Doktora Konsorsiyumu. 17-20 Ekim 2017

