Author Name Disambiguation for the Inspire Project

Mateusz Susik

14 marca 2018

INSPIRE

Welcome to INSPIRE, the High Energy Physics information system. Please direct questions, comments or concerns to feedback@inspirehep.net.

HEP :: HEPNAMES :: INSTITUTIONS :: CONFERENCES :: JOBS :: EXPERIMENTS :: JOURNALS :: HELP

Hawking, Stephen W.

Profile Name | Search

View Profile | Manage Profile | Manage Publications | Help

2015-11-09 14:55:29

PERSONAL INFORMATION

Personal Details (HepNames)

Name	Stephen William Hawking
Current Institution	Cambridge U., DAMTP
E-mails	swh1@damtp.cam.ac.uk Prof.Hawking@damtp.can.ac.uk
Links	http://www.hawking.org.uk https://twitter.com/Prof_S_Haw...
Fields	HEP-TH GR-QC
Identifiers	BAI: S.W.Hawking.1 INSPIRE: INSPIRE-00140145 GoogleScholar: qj74uXkAAAAJ

Period	Rank	Institution
1959 – 1962	UG	Oxford U.
1962 – 1965	PHD	Cambridge U.
1965	SENIOR	Cambridge U., DAMTP

Update Details

Name Variants

Hawking, Stephen W. (11)
Hawking, Stephen (18)
Hawking, S.W. (180)
Hawking, S. W. (1)
Hawking, S. (14)

Affiliations

Cambridge U. (135)
Cambridge U., DAMTP (59)
Caltech (8)
Cambridge U., Inst. of Astron. (4)
Newton Inst. Math. Sci., Cambridge (2)
Santa Barbara, KITP (2)
Munich, Max Planck Inst. (1)
UC, Santa Barbara (1)
CAMBRIDGE U. (1)
Tufts U. (1)

Collaborations

No Collaborations

PUBLICATIONS AND OUTPUT

Publications | Datasets | External

1. The Information Paradox for Black Holes
2. Information Preservation and Weather Forecasting for Black Holes
3. Vector Fields in Holographic Cosmology
4. Quantum Probabilities for Inflation from Holography
5. Accelerated Expansion from Negative Λ
6. The dreams that stuff is made of: The most astounding papers of quantum physics - and how they shook the scientific world
7. Local Observation in Eternal inflation
8. The No-Boundary Measure in the Regime of Eternal Inflation
9. General Relativity
10. General Relativity

Click here to see all

Co-Authors

G.W.Gibbons.1 (14)
T.Hertog.1 (13)
J.B.Hartle.1 (11)
R.Bousso.1 (11)
D.N.Page.1 (8)
C.N.Pope.2 (5)
H.S.Reall.1 (5)
C.Hunter.1 (4)
N.Turok.1 (4)
S.F.Ross.1 (4)
more

Papers

	All papers	Single authored
All papers	224	114
Book	17	7
ConferencePaper	52	44
Introductory	12	9
Lectures	10	10
Published	142	55
Review	5	3
Thesis	0	0
Proceedings	3	0

Subject Categories

Gravitation and Cosmology (128)
Theory-HEP (94)
Astrophysics (55)
General Physics (22)
Phenomenology-HEP (5)
Math and Math Physics (1)
Other (1)

Frequent Keywords

black hole (43)
general relativity (38)
boundary condition (28)
inflation (27)
quantum gravity (27)
cosmological model (19)
path integral (19)
wormhole (19)
field theory: scalar (18)
BOUNDARY CONDITION (16)
more

STATS

Citations Summary

224 papers found, 162 of them citeable (published or arXiv)

	Citeable papers	Published only
Number of papers analyzed:	162	142
Number of citations:	37715	37498
Citations per paper (average):	232.8	264.1
h_{HEP} index [?]	78	78

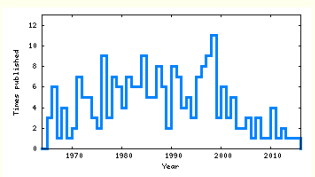
Breakdown of papers by citations:

	Citeable papers	Published only
Renowned papers (500+)	16	16
Famous papers (250-499)	18	18
Very well-known papers (100-249)	32	32
Well-known papers (50-99)	27	26
Known papers (10-49)	41	35
Less known papers (1-9)	20	13
Unknown papers (0)	8	2

Click here to view statistics without self-citations or RPP

Warning: The citations count should be interpreted with great care. Read the fine print

Publication Graph



Jak sobie poradzisz?

Effects of Limited Calorimeter Coverage on ET

Frank E. Paige, A.V. Vanyashin (SSCL)

Mar 1992 - 9 pages

Search for single b^*-quark production with the ATLAS detector at $\sqrt{s} = 7$ TeV

Planck Inst.) , Wainer Vandelli (CERN) , Alexandre Vaniachine (Argonne) , Peter Vankov (DESY) , Francois Vannucci (Paris U., VI-VII) , Riccardo Vari (INFN, Rome) , Erich Varnes (Arizona U.) ,

Jan 2013 - 11 pages

Phys.Lett. B721 (2013) 171-189
(2013-04-25)
DOI: 10.1016/j.physletb.2013.03.016
CERN-PH-EP-2012-344
e-Print: arXiv:1301.1583 [hep-ex] | PDF
Experiment: CERN-LHC-ATLAS

Jak sobie poradzisz?

Effects of Limited Calorimeter Coverage on ET

Frank E. Paige, A.V. Vanyashin (SSCL)

Mar 1992 - 9 pages

Search for single b^*-quark production with the ATLAS detector at $\sqrt{s} = 7$ TeV

Planck Inst.) , Wainer Vandelli (CERN) , Alexandre Vaniachine (Argonne) , Peter Vankov (DESY) , Francois Vannucci (Paris U., VI-VII) , Riccardo Vari (INFN, Rome) , Erich Varnes (Arizona U.) ,

Jan 2013 - 11 pages

Phys.Lett. B721 (2013) 171-189
(2013-04-25)
DOI: 10.1016/j.physletb.2013.03.016
CERN-PH-EP-2012-344
e-Print: arXiv:1301.1583 [hep-ex] | PDF
Experiment: CERN-LHC-ATLAS

✓ Ten sam autor

Homonimy wśród chińskich nazwisk pisanych w języku angielskim

Poznajcie Yanga Wanga, Wanga Yanga i Yanga Wanga!

Definicje


Publications
p_0
p_1
p_{N-1}
Signatures
s_0
s_1
s_2
...
Signature for Doe, John
Title
Lorem ipsum dolor sit amet, consectetur adipiscing elit
Author
Doe, John
Affiliation
University of Foo
Co-authors
Smith, John; Chen, Wang
Year
2015

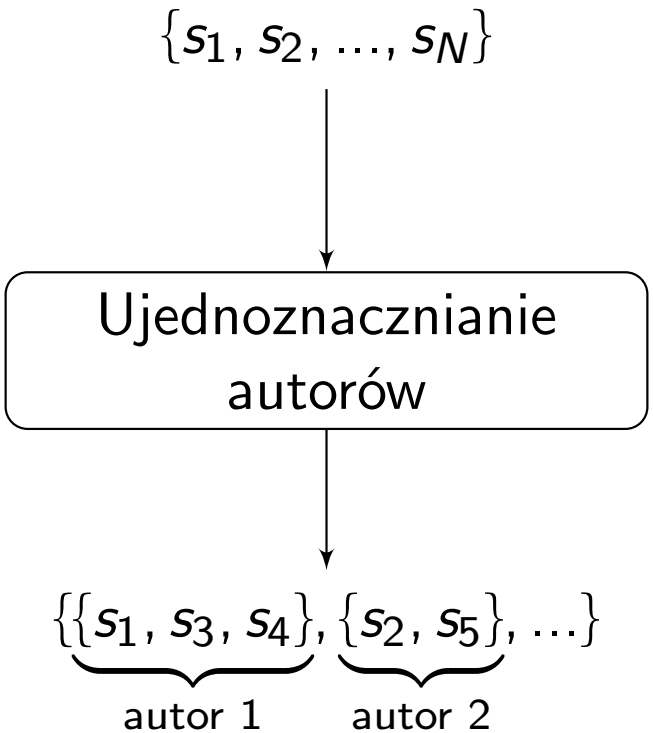
Problem do rozwiązania

Zgrupuj publikacje każdego autora, i tylko jego.

Inspirehep.net jest biblioteką cyfrową, która zawiera:

- 1M publikacji zawierających 10M sygnatur
- 1.2M sygnatur jest oznaczonych:
  - Osobiście przez autorów (podobnie jak w Google Scholar).
  - Za pomocą identyfikatorów (ORCiD).
  - Przez profesjonalnych kustoszy.

Sformułowanie problemu



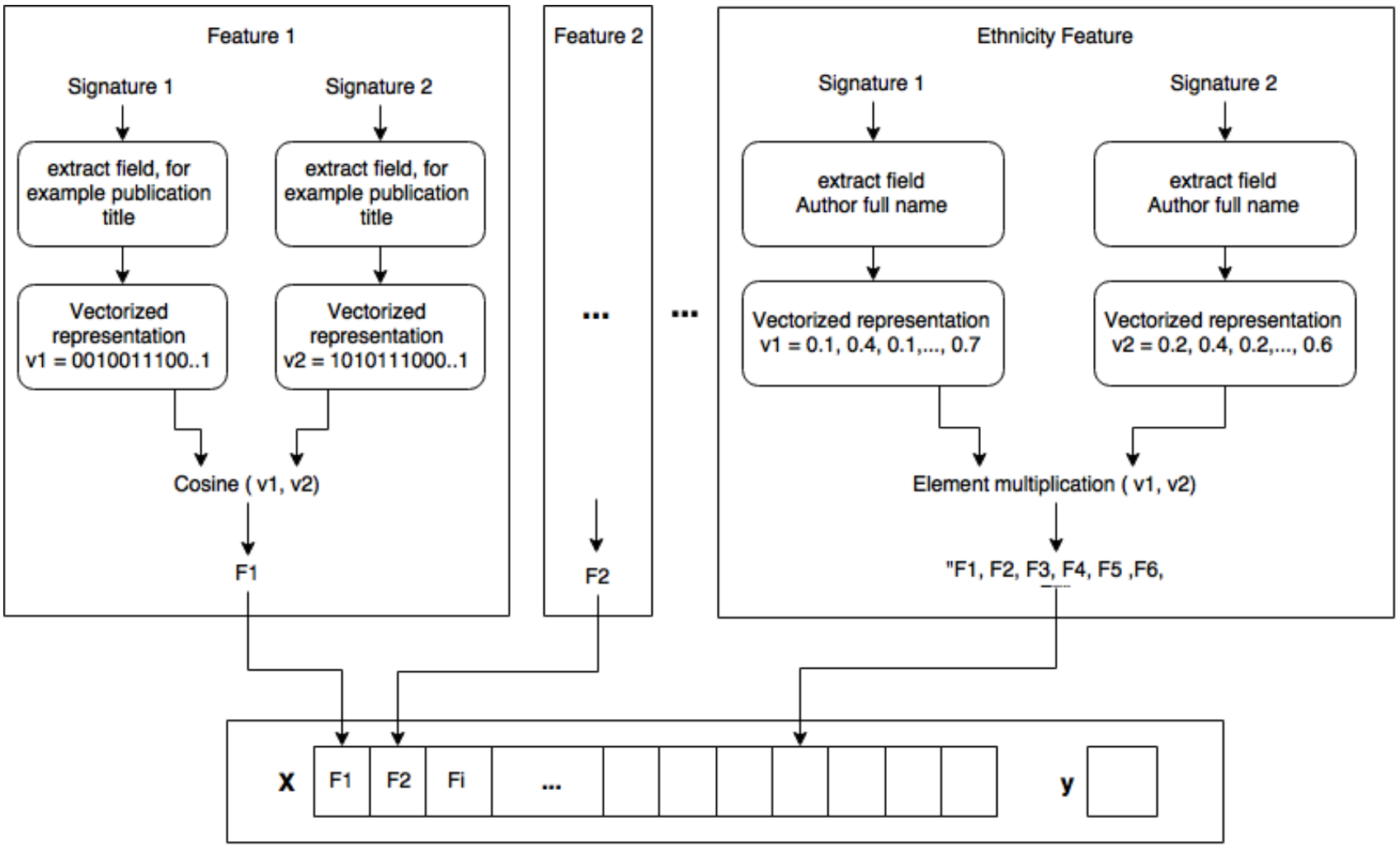
Uczenie się z danych

- Ręczne ujednoznacznianie jest żmudne i trudne, nawet dla doświadczonych kustoszy.
- Czy nie moglibyśmy automatycznie znaleźć zbioru reguł, które ujednoznaczą parę sygnatur?

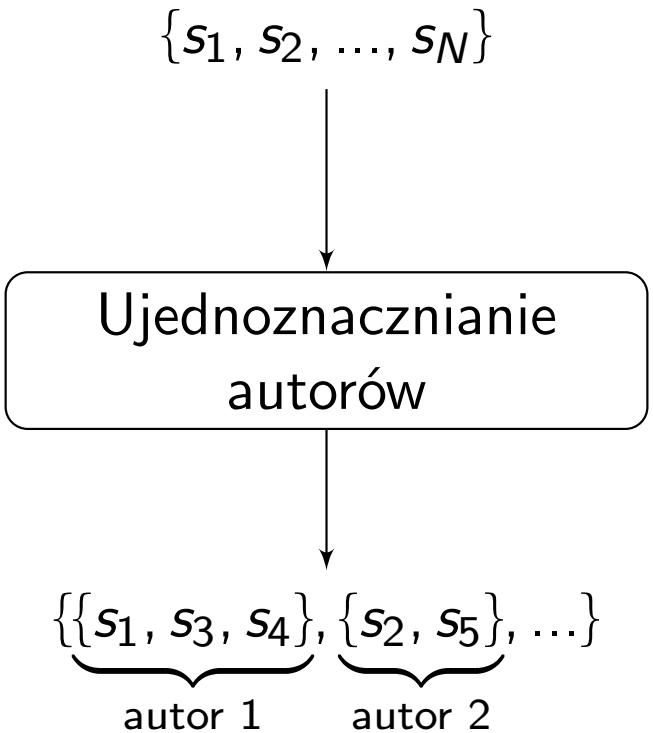
$$\varphi(s_1, s_2) = \begin{cases} 0 & \text{jeśli } s_1 \text{ i } s_2 \text{ należą do tego samego autora,} \\ 1 & \text{wpp.} \end{cases}$$

- Jest to przykład uczenia nadzorowanego.

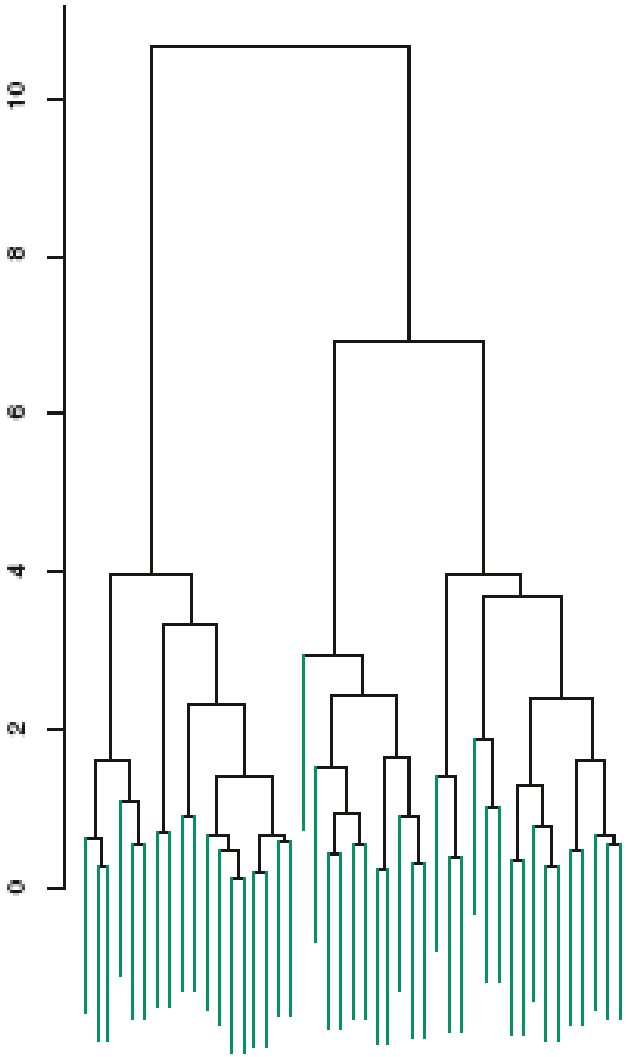
Tworzenie cech na podstawie par sygnatur



Ujednoznacznianie jako problem klastrowania



Aglomeracyjne klastrowanie hierarchiczne

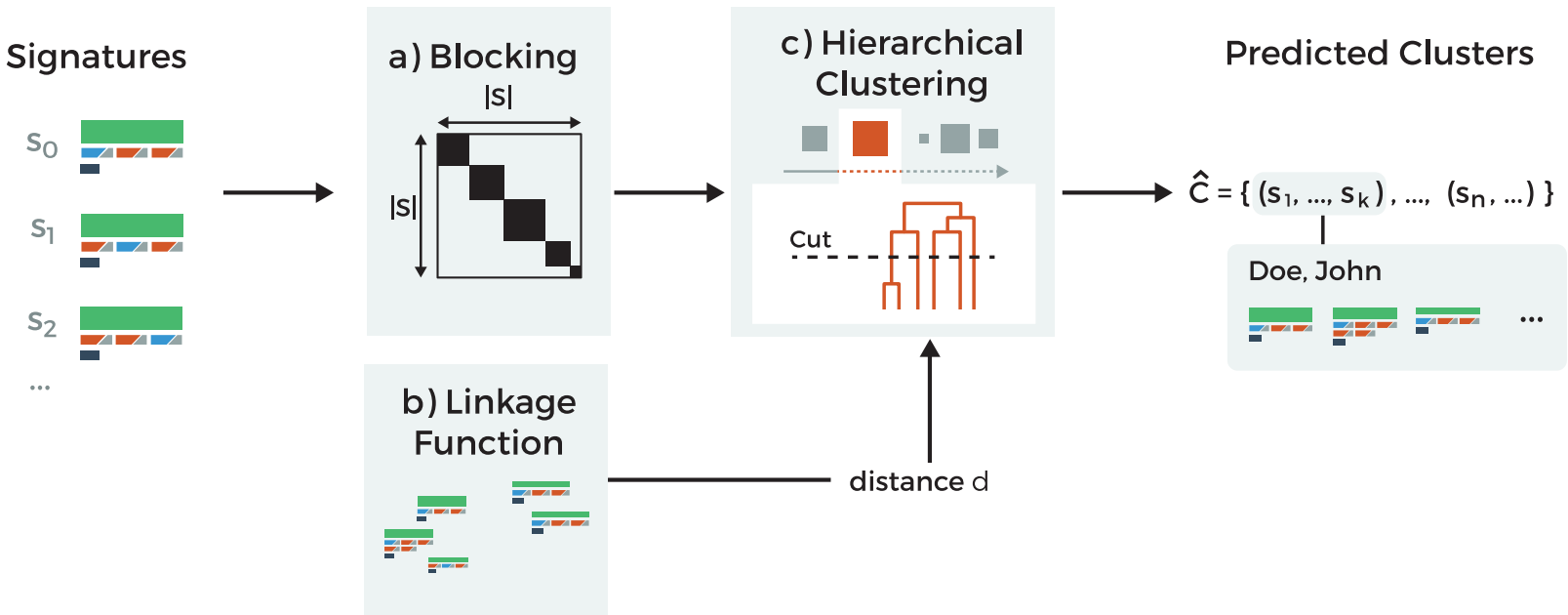


- Rodzina algorytmów klastrujących, która buduje klastry iteracyjnie je łącząc.
- Hierarchia klastrów może być reprezentowana jako dendogram.
- Korzeń drzewa jest klastrem zawierającym wszystkie sygnatury, liście to pojedyncze sygnatury

Problemy z klastrowaniem

- Złożoność tego klastrowania to $O(N^2)$. Mamy $N = 10^7$ sygnatur.
  Rozwiązanie: podzielenie sygnatur na bloki i następnie potraktowanie tych bloków jako oddzielne zagadnienia ujednoznaczniania. Można np. grupować razem sygnatury z tym samym nazwiskiem.
- Jak zdecydować jak wysoko przeciąć dendogram?
  Rozwiązanie: używając oznaczonych sygnatur, wybieramy wysokość, która maksymalizuje wynik B^3 F-measure.

Cały algorytm

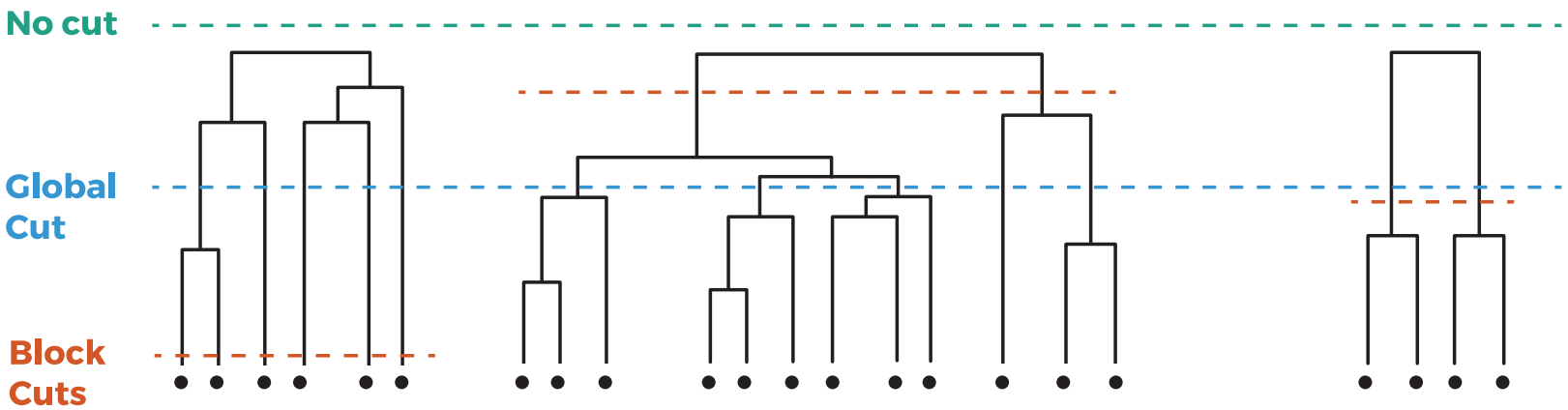


Przypadki

Przeprowadzone analizy pokazują, że jest parę przypadków, gdzie grupowanie na podstawie nazwiska nie działa:

- "Mueller, R." i "Muller, R."
- "Martinez Torres, A." i "Torres, A. Martinez"
- "Smith-Jones, A." i "Smith, A."
- Nazwisko autora zmieniło się (np. przez ślub).

Jak wysoko ciąć dendogramy



Wyniki 1/2

Opis	B^3 *Prec.*	B^3 *Recall*	B^3 *F*1
Prosty algorytm	0.9024	0.9828	0.9409
Grupowanie = Nazwisko i pierwszy inicjał	0.9901	0.9760	0.9830
Grupowanie = Double metaphone	0.9856	0.9827	0.9841
Grupowanie = NYSIIS	0.9875	0.9826	**0.9850**
Grupowanie = Soundex	0.9886	0.9745	0.9815
Klasyfikator = Gradient Boosting DT	0.9901	0.9760	0.9830
Klasyfikator = Lasy losowe	0.9909	0.9783	**0.9846**
Klasyfikator = Regresja Liniowa	0.9749	0.9584	0.9666

Wyniki 2/2

Opis	B^3 *Prec.*	B^3 *Recall*	B^3 *F*1
Prosty algorytm	0.9024	0.9828	0.9409
Klastrowanie = Average linkage	0.9901	0.9760	**0.9830**
Klastrowanie = Single linkage	0.9741	0.9603	0.9671
Klastrowanie = Complete linkage	0.9862	0.9709	0.9785
No cut (prosty algorytm)	0.9024	0.9828	0.9409
Global cut	0.9892	0.9737	0.9814
Block cut	0.9901	0.9760	**0.9830**
Najlepsze ustawienia razem	0.9888	0.9848	**0.9868**

Implementacja

Rozwiązanie jest obecnie używane w CERNie przez projekty INSPIRE i INVENIO.

Czas wykonania: 20 godzin dla 10M sygnatur, na 16 rdzeniach i 32GB RAMu.
Ale ujednoznacznianie inkrementalne zajmuje tylko parę minut!

Nasze rozwiązanie jest open-source[1] i udostępniliśmy nasz zbiór danych[2].

[1] github.com/inspirehep/beard
[2] github.com/glouppe/paper-author-disambiguation/data

Dziękuję za uwagę!