

Les Humanités numériques, ou l'ordinateur nous débarrasse-t-il des livres ?

Frank Fischer

Göttingen Centre for Digital Humanities

frank.fischer@zentr.uni-goettingen.de

Cité Scolaire Gambetta-Carnot, Arras, 21 avril 2015

Les Humanités numériques (Digital Humanities):

- **construire** des ressources numériques
 - **explorer** ces ressources numériques automatiquement (avec des ordinateurs)

La question centrale est toujours:

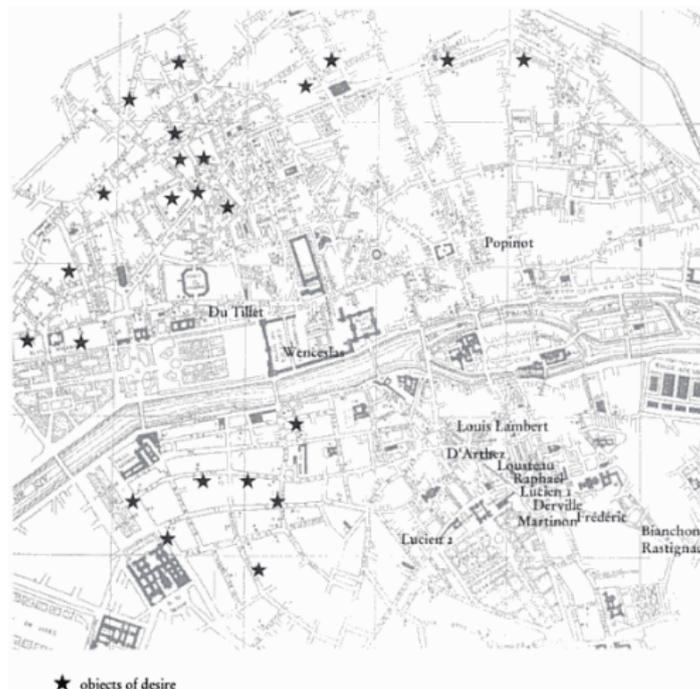
Quel nouveau type de connaissances pouvons-nous
gagner des méthodes numériques ?

Distant Reading

= Lecture à distance

Concept introduit dans l'essai de Franco Moretti, « Conjectures on World Literature », paru dans *New Left Review*, 2000

Distant Reading



Les romans de Balzac : domiciles des protagonistes à Paris

Trois projets poursuivis au centre d'Humanités numériques de l'université de Göttingen:

- ① Un calendrier de la littérature mondiale
- ② Analyse numérique de réseau de pièces de théâtre
- ③ La littérature mondiale selon Wikipédia

Mots-clés

Software et bases de données:

- Android
- Wikipédia
- Java
- Python

Formats et méthodes:

- Analyse des réseaux sociaux
- Web sémantique (Semantic Web)
- traitement automatique du langage naturel (Natural-Language Processing)

- XML

Entités littéraires:

- Distant Reading
- Pièces de théâtre
- Prose
- Études de réception
- Questions de corpus
- Littérature mondiale (de, fr, en, es, ru, ...)

Suomen kirjallisuus

« *La plupart des romans finlandais ont lieu **en été.*** »

Entretien avec le traducteur Stefan Moster,
Deutschlandradio Kultur, 8 octobre 2014

Bloomsday

« *Miss Dunne clicked on the keyboard:*
— *16 June 1904.* »

James Joyce: *Ulysses* (publ. 1922)

Dans la montagne

« *Den 20. Jänner ging Lenz durchs Gebirg.* »

Georg Büchner: *Lenz* (publ. 1839)

Sorties avec maman

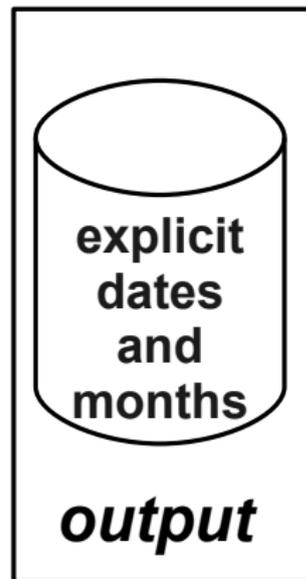
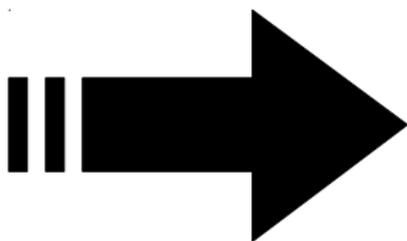
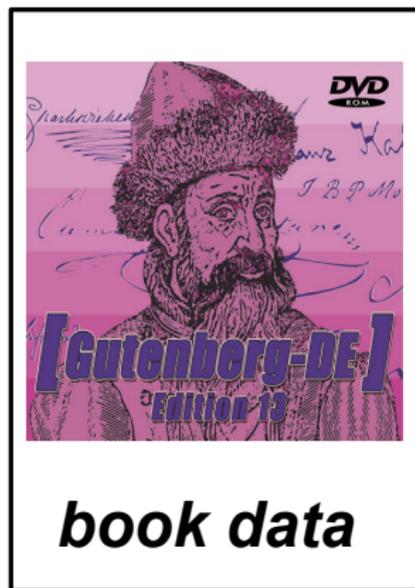
« *Quand vint le 1er janvier, je fis d'abord des visites de famille avec maman, (...)* »

Proust: *À l'ombre des jeunes filles en fleurs* (publ. 1919)

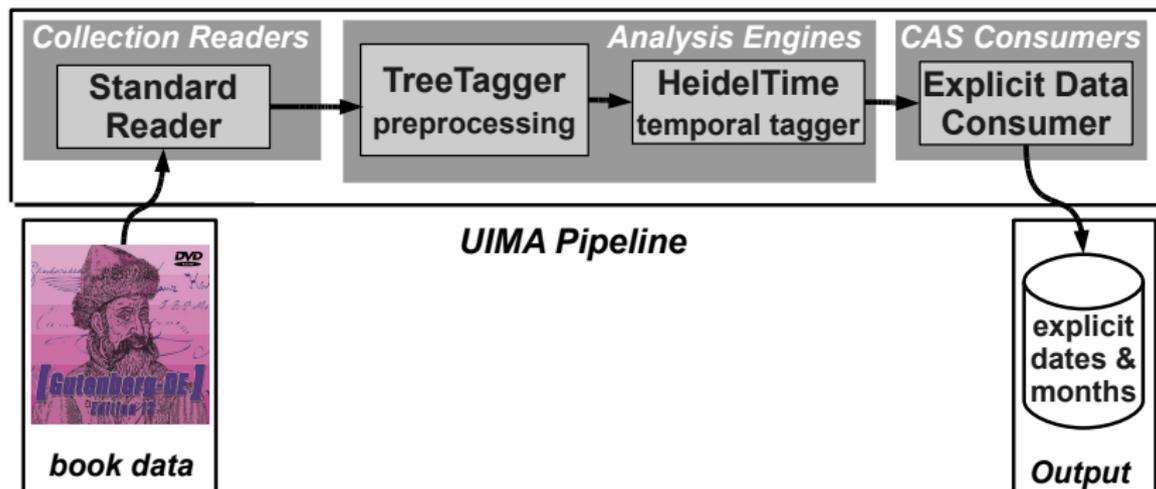
Gutenberg-DE

- Source: Gutenberg-DE, 13ème édition, DVD-ROM (Octobre 2013)
- 891 Mo de romans et nouvelles en langue allemande
- en tout, 549 écrivains et 2.735 œuvres de 1510 aux années 1940

Extraction d'indications normalisées de jours et de mois

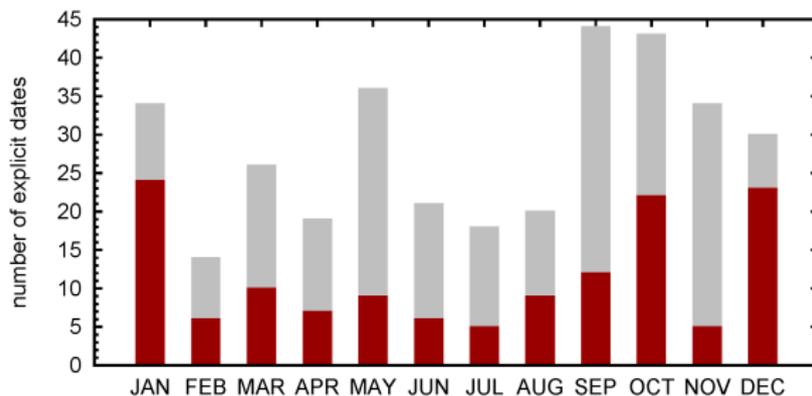


Extraction d'indications normalisées de jours et de mois



Ce qui sort du pipeline ...

“Mois vs. jours”



Ce qui sort du pipeline ...

Dates explicites en cartes thermiques :

```

EXPLICIT DAYS: 81
MONTH: 01 21321111---2-211-121-1--1-----
MONTH: 02 1-11-----1--1--1-----
MONTH: 03 -----2---1---2-----2--1-2--
MONTH: 04 1-----1-2-----1-----
MONTH: 05 321--2----1-----
MONTH: 06 2-----1----1--1-1-
MONTH: 07 2-1-----1-----1---
MONTH: 08 -----12-2112-----
MONTH: 09 -3---1-1--2-----3----1--1
MONTH: 10 7-82-----4-----1-----
MONTH: 11 -----1--12-----1---
MONTH: 12 -6-1--1--12-----1--4-1--123
  
```

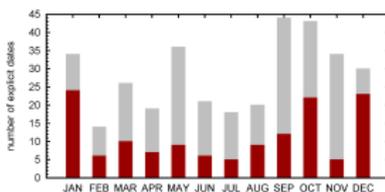
Ce qui sort du pipeline ...

Extraits avec dates explicites :

01-01	Die Poggenpuhls (1896)	1. Januar	SATZ: Und so vernimm denn, am Silversterabend noch ein Bettler (allerdings ein glücklicher, denn wir brachten es im Kasino auf sieben Bowlen in Großformat) und am 1. Januar früh ein Gott, ein Krösus.
01-01	Mathilde Möhring (1908)	erste Januar	SATZ: Der erste Januar war ein wundervoller Wintertag, alles überreift und übereist, aber nicht sehr kalt und eine helle Sonne am blauen Himmel.
01-02	Meine Kinderjahre (1894)	zweiten Januar	SATZ: Silvester war wie herkömmlich gefeiert worden, und für den zweiten Januar stand ein neues Vergnügen in Sicht, von dem ich mir ganz besonders viel versprach:
01-03	Die Poggenpuhls (1896)	dritte Januar	SATZ: Es war ein Wintertag, der dritte Januar.
01-03	Effi Briest (1895)	3. Januar	SATZ: Seit Silvesternacht ging ein scharfer Nordost, der sich in den nächsten Tagen fast bis zum Sturm steigerte, und am 3. Januar nachmittags hieß es, daß ein Schiff draußen mit der Einfahrt nicht zustande gekommen und hundert Schritt vor der Mole gescheitert sei;
01-03	Meine Kinderjahre (1894)	dritten Januar	SATZ: mein Freund Wilhelm Krause, der schon als Schüler und Pensionär des bekannten Direktors v. Klöden die Gewerbeschule besuchte, mußte am dritten Januar wieder in Berlin sein, und seitens seines Vaters, des Kommerzienrats, war mit einigen Freunden verabredet worden, dem liebenswürdigen Jungen bis nach dem jenseitigen Ufer hinüber, von wo dann die Fahrpost ging, das Geleit zu geben.

Ce qui sort du pipeline ...

Mois vs. jours



Cartes thermiques

```

EXPLICIT DAYS: 81
MONTH: 01 21321111--2-211-121-1-1-----
MONTH: 02 1-11-----1-1-1-----
MONTH: 03 -----2-1-1---2-----2-1-2--
MONTH: 04 1-----1-2-----1-----
MONTH: 05 321-2-----1-----
MONTH: 06 2-----1-----1-----1-1-1-
MONTH: 07 2-1-----1-----1-----
MONTH: 08 -----12-2112-----
MONTH: 09 -3---1-1-1-2-----3---1-1-
MONTH: 10 7-82-----4-----1-----
MONTH: 11 ---1-12-----1-----1-----
MONTH: 12 -6-1-1-12-----1-4-1-123
  
```

Dates explicites

```

[1-01] Sie Poggenpohl (1896) 1. Januar
SATZ: Und so verriech denn, an Silvesterabend noch
ein bisschen Gulliver's Reise gleichfalls, denn wir
brachten es im Kalise auf sieben Bänden in
deutscher) und am 1. Januar FROB ein Satz, ein
Kreuz.
[01-01] Mathias Möhring (2008) erste Januar
SATZ: Der erste Januar war ein sonderbarer
Mistertag, alles überflutet und überweht, aber nicht
eher kalt und eine helle Sonne am blauen Himmel.
[01-02] Meine Kinderjahre (1884) zweiten Januar
SATZ: Silvester war wie herkömmlich gefeiert worden,
und für den zweiten Januar stand mir neues Übergangs
in Sicht, von dem ich mir ganz besonders viel
versprach.
[01-03] Sie Poggenpohl (1896) dritte Januar
SATZ: In der aus Mitternacht, der dritte Januar.
[01-03] Effi Briest (1895) 3. Januar
SATZ: Seit Silvesterabend ging ein sicherer Herbst,
der mich in dem nächsten Tagen fast bis zum Stern
steigerter, und am 3. Januar nachmittags ließ es, daß
die Welt endlich mit der Einnahme nicht zulaufen
gekommen und Hundert Schritt vor der Welt geschickter
blieb.
[01-03] Meine Kinderjahre (1884) dritten Januar
SATZ: mein Freund Wilhelm Krause, der schon als
Schüler und Ausbilder des bekannten Bruckers v.
Klause die amerikanische Beweise, hatte am dritten
Januar wieder in Berlin sein, und seinen seinen
Vater, des Kommerzienrats, war mit einigen Freunden
verabschiedet worden, des siebenzehnjährigen Jungen bis
nach dem jeweiligen der Handler, von wo denn die
Einkauf ging, den Gehalt zu geben.
  
```

alle Statistiken pro Werk, Autor, Epoche, ...

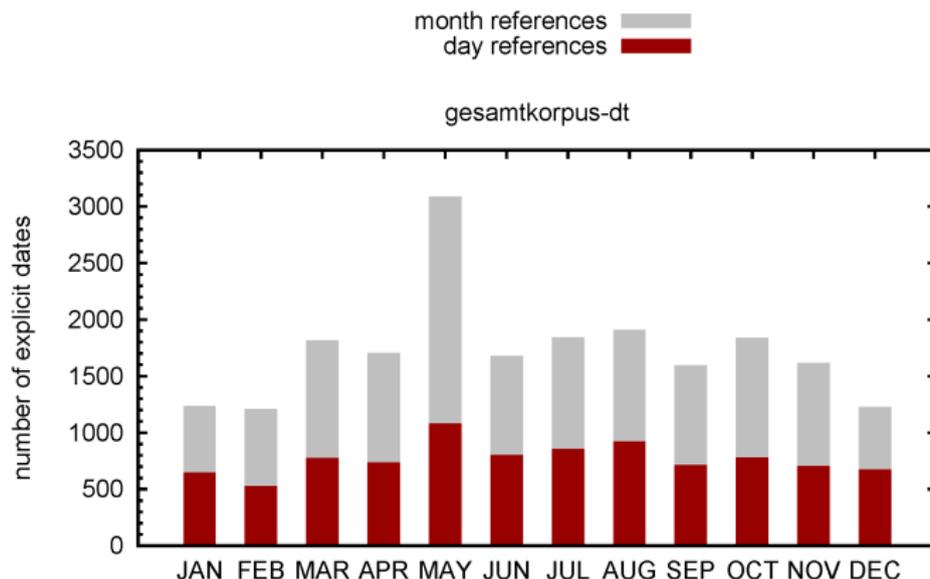
Au mois de mai

*« Au mois de mai, quand la lumière
Voyait tous les bourgeons s'ouvrir,
L'amour, en sa douceur première,
Dans mon cœur s'est mis à fleurir.*

*Au mois de mai, sous la rainée,
Tous les oiseaux chantaient en chœur
Quand j'ai dit à la bien-aimée
Le tendre secret de mon cœur. »*

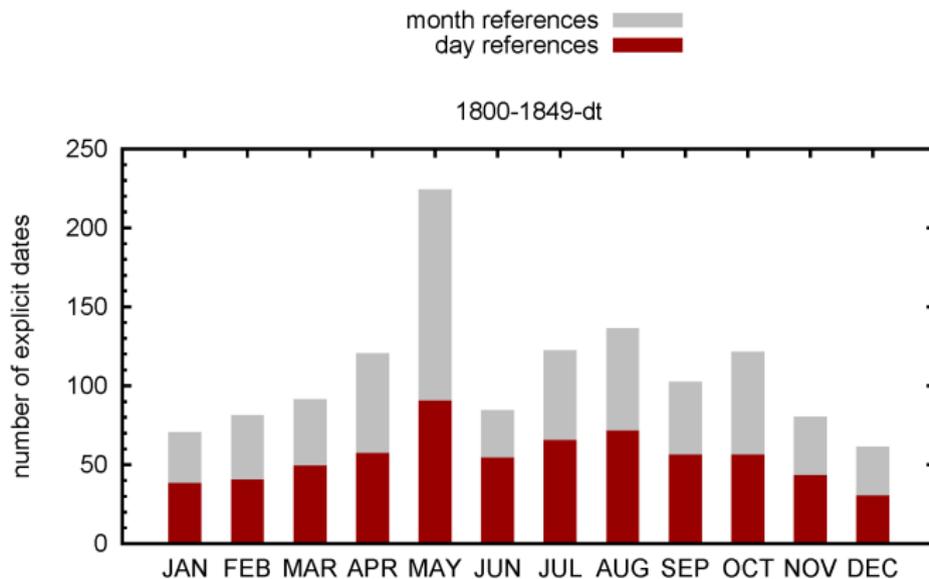
Heinrich Heine: *Buch der Lieder* (publ. 1827)

Indications de jours et de mois



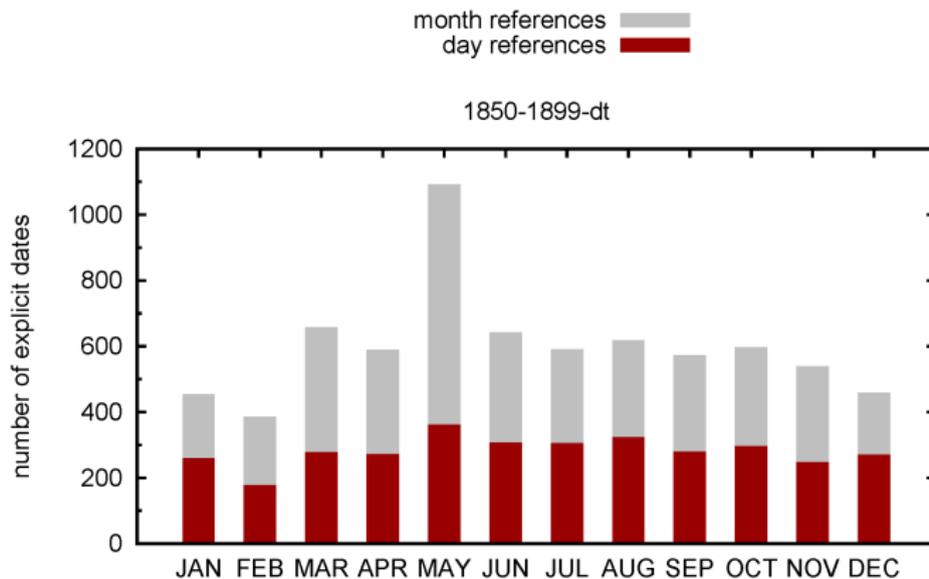
Corpus intégral de la littérature allemande : claire victoire de mai, résultats honorables pour le printemps et l'été, l'automne en difficulté, les mois de novembre à février sont battus à plate couture.

Corpus partiel de 1800 à 1849



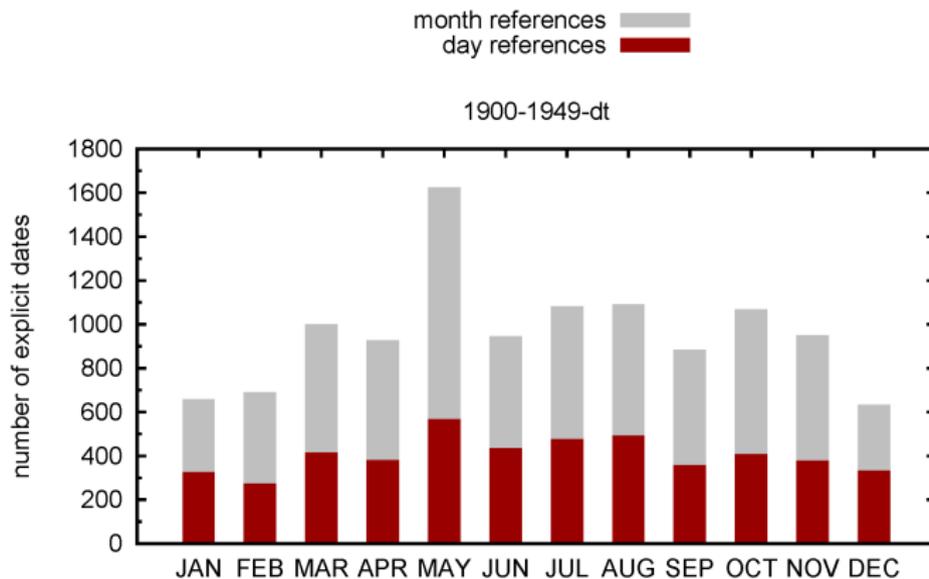
Et encore le mois de mai !

Corpus partiel de 1850 à 1899



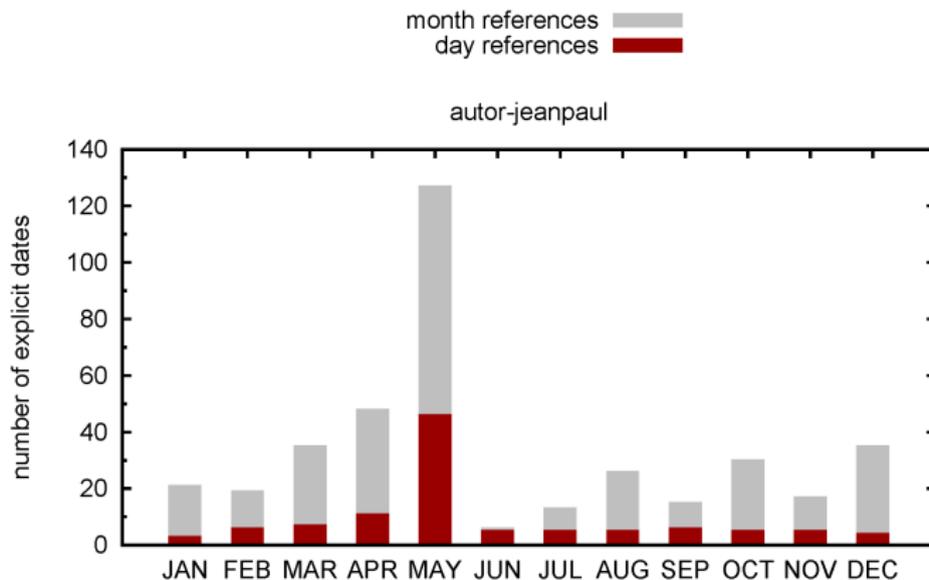
Et encore !

Corpus partiel de 1900 à 1949



Et encore !

Jean Paul



Jean Paul, le plus grand amateur du mois de mai !

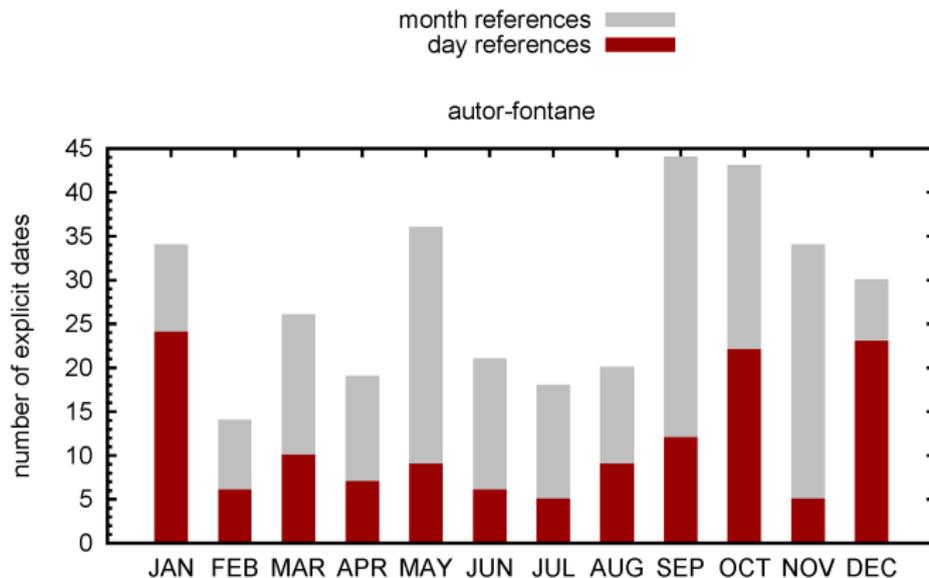
Quand est-ce que a lieu *Effi Briest* ?



R. W. Fassbinder: *Fontane Effi Briest* (1974)

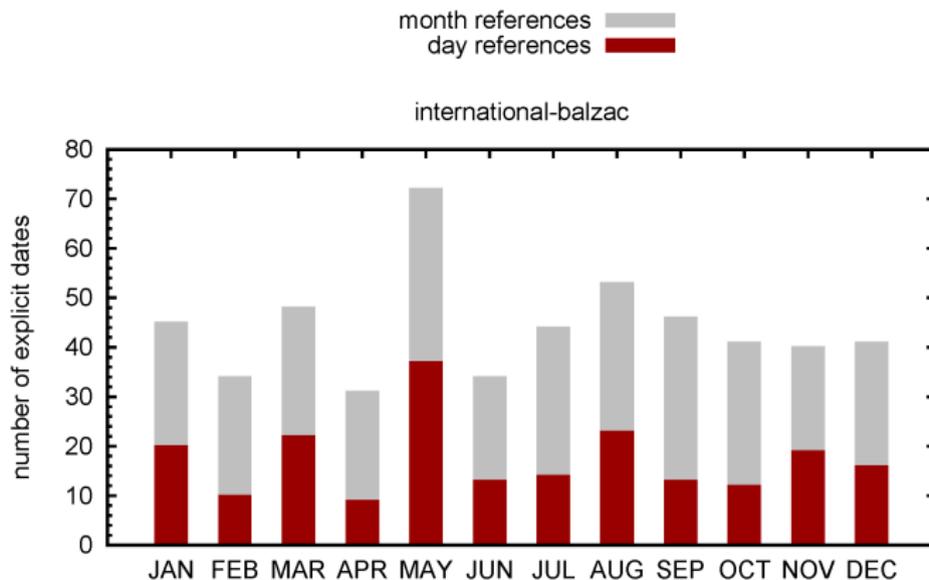
Captures d'écran © Basis-Film Verleih

Theodor Fontane



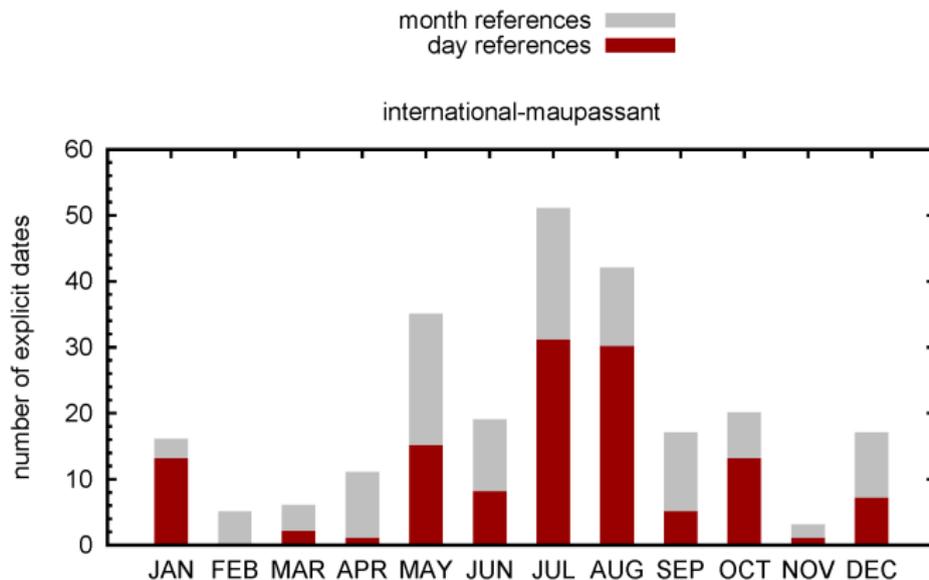
Theodor Fontane, auteur de l'automne et de l'hiver !

Balzac



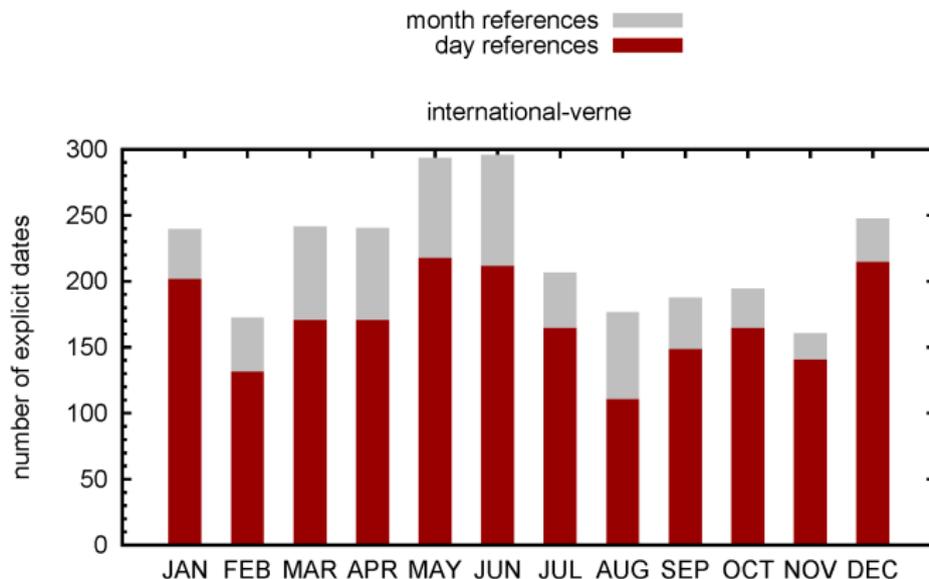
Balzac aussi préfère le mois de mai !

Guy de Maupassant



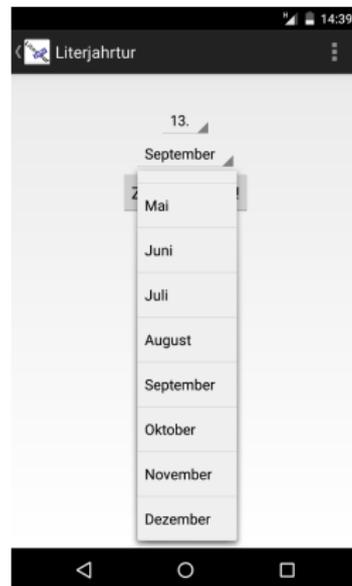
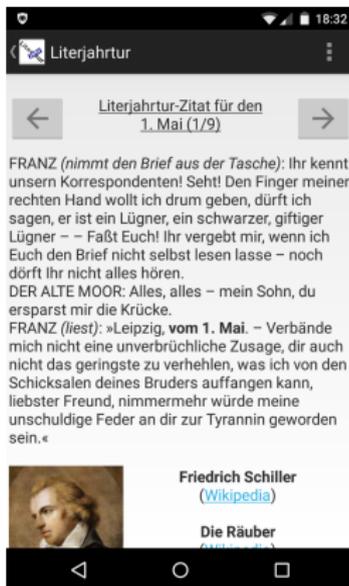
Maupassant, homme de l'été!

Jules Verne, le “calendar boy” de la littérature mondiale



Les romans d'aventures de Jules Verne ont toujours une chronologie très précise.

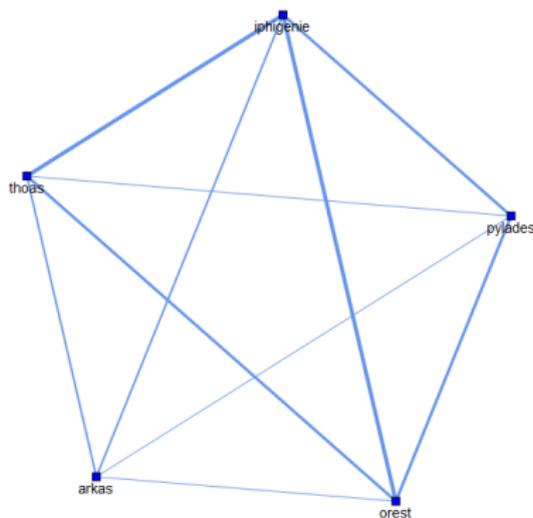
Application pour Android



Télécharger la version Beta dans Google Play Store :

<https://goo.gl/rF4Nbj>

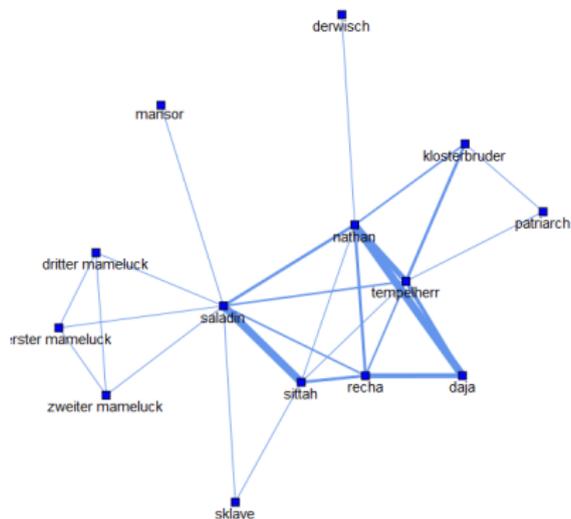
Graphe du réseau de « Iphigénie en Tauride » (1787)



Goethe: Iphigenie auf Tauris (1787) – Density: 1.0

J. W. von Goethe: *Iphigénie en Tauride* (1787)

Graphe du réseau de « Nathan le Sage » (1779)



Lessing: Nathan der Weise (1779) – Density: 0.3

G. E. Lessing: *Nathan le Sage* (1779)

Valeurs de centralité

G. E. Lessing: Nathan le Sage (1779)

Nombre de personnages: 14

Degré maximal: 9

Degré moyen: 3.86

Densité: 0.3

Longueur moyen de chemin: 1.93

Personnage	Centralité d'intermediarité	Degré	Moyenne de distance	Centralité de proximité
daja	0.0	3	2.15	0.46
nathan	18.83	7	1.46	0.68
recha	2.33	5	1.62	0.62
tempelherr	16.83	7	1.46	0.68
saladin	42.5	9	1.31	0.76
sittah	3.5	5	1.62	0.62
derwisch	0.0	1	2.38	0.42
klosterbruder	1.0	3	2.15	0.46
patriarch	0.0	2	2.31	0.43
sklave	0.0	2	2.15	0.46
zweiter mameluck	0.0	3	2.08	0.48
erster mameluck	0.0	3	2.08	0.48
dritter mameluck	0.0	3	2.08	0.48
mansor	0.0	1	2.23	0.45

Centralité d'intermediarité

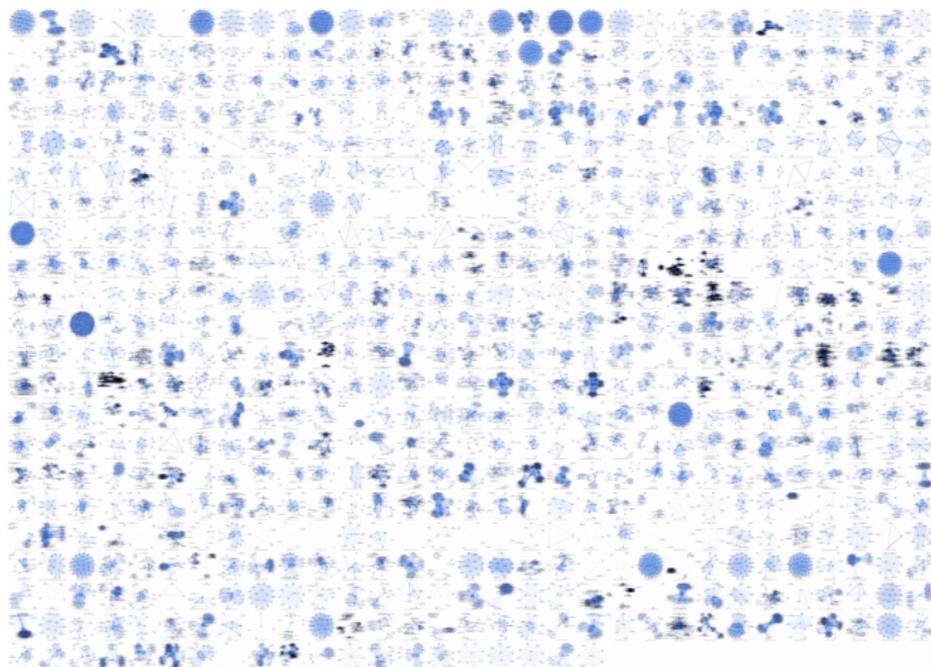
The betweenness centrality of a node v is given by the expression:

$$g(v) = \sum_{s \neq v \neq t} \frac{\sigma_{st}(v)}{\sigma_{st}}$$

where σ_{st} is the total number of shortest paths from node s to node t and $\sigma_{st}(v)$ is the number of those paths that pass through v .

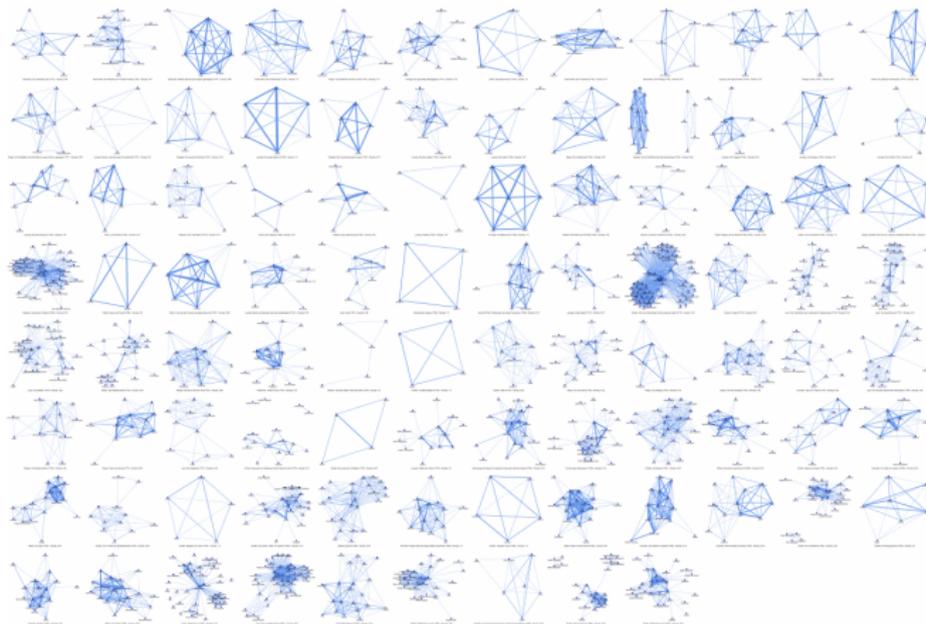
Betweenness Centrality: Voilà la formule.

Graphes de réseau de 671 pièces de théâtre



671 pièces de théâtre allemands

93 pièces du 18ème siècle



93 pièces de théâtre allemandes (1731–1799)

Caractéristiques des 15 plus grandes versions de Wikipedia

Langue	Code	#Articles	Depth	#Edits	#Écrivains	#Écrivains dans notre set
English	en	4,738,257	895.18	759,894,686	25,995	10,765
German	de	1,822,322	88.32	145,142,072	0	5,202
French	fr	1,598,081	196.93	114,926,655	7,974	4,981
Russian	ru	1,197,015	116.73	81,444,167	11,515	3,286
Italian	it	1,177,497	107.40	76,520,585	0	3,514
Spanish	es	1,161,825	177.01	85,947,882	6,668	3,430
Portuguese	pt	867,553	124.47	42,572,971	1,104	2,226
Chinese	zh	814,358	134.23	35,959,378	–	1,468
Persian	fa	446,641	209.27	19,684,602	–	1,395
Arabic	ar	352,720	202.99	17,792,878	1,288	1,017
Korean	ko	306,641	85.53	15,571,911	968	1,052
Hungarian	hu	286,466	91.84	16,382,712	3,033	1,149
Serbo-Croatian	sh	278,382	354.65	9,143,222	–	849
Romanian	ro	271,118	93.13	9,489,089	–	1,617
Turkish	tr	242,778	205.94	16,123,456	1,314	1,217

Infobox de Wikipedia

William Shakespeare



The Chandos portrait, artist and authenticity unconfirmed. National Portrait Gallery, London.

Born	Baptised 26 April 1564 (birth date unknown) Stratford-upon-Avon, Warwickshire, England
Died	23 April 1616 (aged 52) Stratford-upon-Avon, Warwickshire, England
Occupation	Playwright, poet, actor
Period	English Renaissance
Spouse	Anne Hathaway (m. 1582-1616)
Children	Susanna Hall Hamnet Shakespeare Judith Quiney
Relatives	John Shakespeare (father) Mary Shakespeare (mother)
Signature	

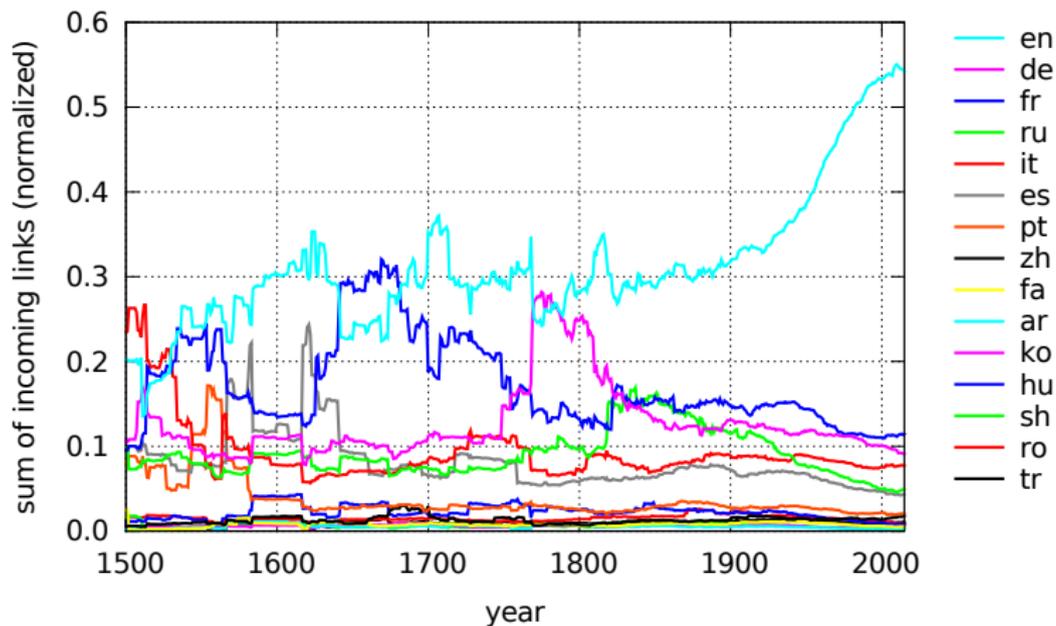
Voltaire



Portrait de Voltaire d'après Quentin de La Tour. (vers 1736)
Château de Ferney-Voltaire

Nom de naissance	François-Marie Arouet
Alias	Voltaire
Naissance	21 novembre 1694 Paris (France) Royaume de France
Décès	30 mai 1778 (à 83 ans) Paris
Activité principale	Ecrivain et philosophe
Auteur	
Langue d'écriture	Français
Mouvement	Philosophe des Lumières
Genres	Roman, conte, théâtre, essai
Œuvres principales	<ul style="list-style-type: none"> • <i>Candide ou l'Optimisme</i> (1759) • <i>L'Ingénu</i> (1767) • <i>Zadig ou la Destinée</i> (1748) • <i>Lettres philosophiques</i> (1734) • <i>Dictionnaire philosophique</i> (1764)
Compléments	<ul style="list-style-type: none"> • Intervenu dans différentes affaires judiciaires dont l'affaire Calas (1763), l'affaire Sirey (1764) et l'affaire du Chevalier de La Barre (1766) • Entré au Panthéon en 1791
modifier	

Nombre de liens (qui renvoient à des articles sur des écrivains) par version Wikipédia



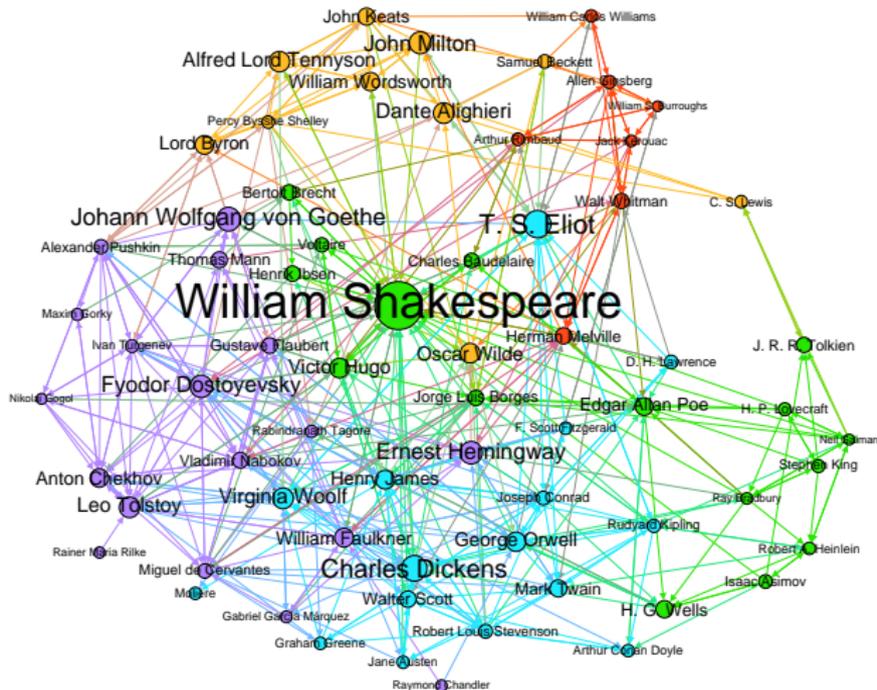
Les écrivains du top 5 (Nombre de liens)

lang	1st	2nd	3rd	4th	5th
ENG	W. Shakespeare	Robert Christgau	Roger Ebert	Charles Dickens	J. R. R. Tolkien
GER	J. W. von Goethe	W. Shakespeare	Friedrich Schiller	Bertolt Brecht	Thomas Mann
FRE	W. Shakespeare	Victor Hugo	Molière	Voltaire	Anton Chekhov
RUS	Alexander Pushkin	W. Shakespeare	Anton Chekhov	Maxim Gorky	Leo Tolstoy
ITA	Dante Alighieri	W. Shakespeare	Cicero	J. R. R. Tolkien	Virgil
SPA	W. Shakespeare	J. R. R. Tolkien	Cervantes	Lope de Vega	Cicero
POR	W. Shakespeare	Robert Christgau	J. R. R. Tolkien	Vinicius de Moraes	Cicero
CHI	Lu Xun	J. R. R. Tolkien	W. Shakespeare	Jin Yong	Li Bai
PER	Ferdowsi	Ahmad Shamloo	W. Shakespeare	Ali-Akbar Dehkhoda	Moh.-Taqi Bahar
ARA	W. Shakespeare	Naguib Mahfouz	J. W. von Goethe	Al-Maqrizi	Ahmed Shawqi
KOR	W. Shakespeare	Yi Kwang-su	J. W. von Goethe	J. R. R. Tolkien	Dante Alighieri
HUN	W. Shakespeare	Sándor Petőfi	János Arany	Attila József	Robert Christgau
SER	W. Shakespeare	Vjekoslav Klišić	Cicero	Roger Ebert	Plutarch
ROM	Ion Luca Caragiale	Mihai Eminescu	W. Shakespeare	Jules Verne	Mircea Eliade
TUR	W. Shakespeare	Bertolt Brecht	J. R. R. Tolkien	Anton Chekhov	Molière

Les écrivains du top 5 (selon l'algorithme PageRank)

lang	1st	2nd	3rd	4th	5th
ENG	W. Shakespeare	T. S. Eliot	Charles Dickens	J. W. von Goethe	Ernest Hemingway
GER	W. Shakespeare	J. W. von Goethe	Voltaire	Thomas Mann	Friedrich Schiller
FRE	W. Shakespeare	Victor Hugo	André Gide	J. W. von Goethe	Charles Baudelaire
RUS	Alexander Pushkin	Heinrich Heine	Alexander Pope	W. Shakespeare	Gavrila Derzhavin
ITA	W. Shakespeare	Dante Alighieri	J. W. von Goethe	Virgil	Petrarch
SPA	W. Shakespeare	Lord Byron	Dante Alighieri	Jorge Luis Borges	Edgar Allan Poe
POR	W. Shakespeare	Stephen King	T. S. Eliot	F. Scott Fitzgerald	Zelda Fitzgerald
CHI	W. Shakespeare	J. W. von Goethe	Cervantes	Charles Baudelaire	Molière
PER	Sanai	W. Shakespeare	Alex. Pushkin	Nikolai Gogol	Ferdowsi
ARA	W. Shakespeare	Christ. Marlowe	Leo Tolstoy	Friedrich Schiller	F. Dostoyevsky
KOR	W. Shakespeare	Dante Alighieri	J. W. von Goethe	Friedrich Schiller	Virgil
HUN	W. Shakespeare	Giovanni Boccaccio	Petrarch	J. W. von Goethe	André Breton
SER	W. Shakespeare	Fyodor Dostoyevsky	Charles Dickens	Virgil	Dante Alighieri
ROM	W. Shakespeare	J. W. von Goethe	Cervantes	Charles Baudelaire	Molière
TUR	W. Shakespeare	Peyami Safa	J. R. R. Tolkien	C. S. Lewis	Charles Baudelaire

Les écrivains du top 60 (dans un graphe en cluster dans la version anglaise de Wikipédia)



Merci !



Des questions ? Frank Fischer <frank.fischer@zentr.uni-goettingen.de>