



HAL
open science

vHALo : Analyse des collaborations entre chercheurs au travers de l'archive ouverte HAL

Emmanuel Giguët

► **To cite this version:**

Emmanuel Giguët. vHALo : Analyse des collaborations entre chercheurs au travers de l'archive ouverte HAL. 2019. hal-02380362

HAL Id: hal-02380362

<https://hal.science/hal-02380362>

Preprint submitted on 10 Aug 2022

HAL is a multi-disciplinary open access archive for the deposit and dissemination of scientific research documents, whether they are published or not. The documents may come from teaching and research institutions in France or abroad, or from public or private research centers.

L'archive ouverte pluridisciplinaire **HAL**, est destinée au dépôt et à la diffusion de documents scientifiques de niveau recherche, publiés ou non, émanant des établissements d'enseignement et de recherche français ou étrangers, des laboratoires publics ou privés.

Le projet vHALo : Analyse des collaborations entre chercheurs au travers de l'archive ouverte HAL

Emmanuel Giguët

Normandie Univ, UNICAEN,
ENSICAEN, CNRS, GREYC
14000 Caen, France
emmanuel.giguët@unicaen.fr

Contexte

En 2018, le CNRS signe la déclaration de San Francisco DORA sur l'évaluation de la recherche qui enjoint de préférer une évaluation qualitative. Initiée par des scientifiques et un groupe d'éditeurs de journaux scientifiques, la déclaration DORA – San Francisco Declaration on Research Assessment – met en cause l'usage croissant du classement bibliométrique, l'Impact Factor, comme indice au service de l'évaluation de la recherche et des chercheurs. Le projet vHALo, aujourd'hui nommé Beyond Rankings, est une contribution au service de l'évaluation qualitative de la recherche basée sur l'archive ouverte HAL.

Résumé

vHALo exploite les collections de publications scientifiques référencées dans HAL pour représenter les collaborations scientifiques et fournir automatiquement une mesure de la qualité des publications. Au delà des indicateurs bibliométriques classiques, vHALo apporte une aide au pilotage de structures, en fournissant des indicateurs qualitatifs, ainsi qu'une aide au montage de projet, en facilitant l'identification des compétences.

Sur le plan scientifique, le projet vHALo permet d'aborder des thématiques diverses :

- en Traitement Automatique des Langues : analyse sémantique des résumés de articles, désambiguïsation des auteurs/structures, enrichissement des informations, analyse des thématiques, classification automatique
- en Visualisation : représentation des connaissances d'interactions, évolutions temporelles, changement d'échelle, ergonomie cognitive
- en Analyse de graphe : analyse de réseaux sociaux, graphes de terrain, graphe « petit-monde », détection de communautés, planarité

Enjeux

En intelligence économique territoriale, la connaissance des acteurs de son réseau, à savoir leur domaine d'expertise, leur orientation, les relations qu'ils entretiennent et développent au niveau local, régional, national ou internationale, est une des clés permettant une analyse suivie de prises de décisions motivées.

L'activité des chercheurs et de leur laboratoire n'échappe pas à ce type de lecture. Elle est l'objet d'observations par de multiples acteurs : qu'il s'agisse d'instances d'évaluation de la recherche, comme l'HCERES, le CNU, ou des instances locales type ACS ; qu'il s'agisse de structures de valorisation de la recherche, de l'innovation ou des relations industrielles, qui souhaitent développer les collaborations institutionnelles et les partenariats industriels ; qu'il s'agisse d'organismes de soutien à la recherche, comme l'ANR ou la BPI.

Le pilotage de structures de recherche, que ce soit par le biais de tutelles ou par les organes de direction des unités de recherche, nécessite également une bonne connaissance des acteurs et de leur environnement. Cela permet notamment de mieux accompagner et valoriser le travail des chercheurs, des équipes, et au final de l'unité de recherche toute entière.

Pour alimenter la réflexion et la stratégie, force est de constater qu'il convient de disposer d'indicateurs témoignant de l'activité de recherche et permettant d'en mesurer la qualité, le plus objectivement possible.

L'analyse de l'activité ne peut se résumer à un bilan quantitatif de la production : nombre de publications, nombre de projets acceptés, nombre de brevets... Il convient de fournir un éclairage plus fin, montrant notamment l'insertion du chercheur, de l'équipe, de l'unité dans son environnement, en caractérisant le dynamisme des relations institutionnelles et industrielles, en exhibant les collaborations, qu'elles soient pluri-disciplinaires ou non, en positionnant l'unité sur la scène locale, régionale, nationale et internationale.

L'étude des thématiques de recherche et des domaines d'application fournit un éclairage complémentaire qui permet également de comprendre l'insertion du chercheur ou de l'unité de recherche dans son environnement, son positionnement par rapport à des unités comparables, que ce soit pour faciliter le montage de projets ANR, pour accompagner un industriel qui cherche des compétences mobilisables (contrat, CIFRE), pour alimenter un GDR, pour identifier les verrous que la communauté traite ainsi que les thématiques en essor.

L'archive ouverte HAL

L'archive ouverte HAL destinée au dépôt et à la diffusion d'articles scientifiques s'impose aujourd'hui comme un des socles opérationnels du plan national pour la Science Ouverte. Gérée par le CCSD, l'archive HAL reçoit un soutien affiché d'opérateurs de recherche dont le CNRS qui impose dorénavant l'utilisation de HAL pour recenser les travaux de recherche.

Il s'agit donc d'une source particulièrement pertinente pour commencer à asseoir une analyse des collaborations entre chercheurs. À terme, des sources ouvertes complémentaires telles que DBLP pourront être connectées au système afin d'étendre le périmètre, notamment à l'international.

Étude des collaborations scientifiques

La visualisation des collaborations entre chercheurs est une source d'information particulièrement intéressante. Ces représentations, produites automatiquement, peuvent être réalisées à différentes échelles – chercheurs, équipes, laboratoires, fédérations, tutelles, pour produire des vues synthétiques adaptées.

Nous avons par exemple montré qu'il était possible d'utiliser HAL pour mettre en évidence les collaborations internes, ainsi que les relations entretenues sur la scène locale, régionale, nationale et internationale. Ces représentations synthétiques ont révélé des formes de collaboration méconnues, voire insoupçonnées.

Les graphes de relations exhibant les relations internes sur une période donnée permettent de caractériser les relations entre chercheurs d'une même équipe, d'identifier les chercheurs les plus isolés ainsi que les fonctionnements en mode projet. Les graphes de relations internes peuvent également se révéler utiles au pilotage en mesurant l'impact social du départ de permanents, quelle qu'en soit la raison : retraite, mobilité, disponibilité, mise à disposition.

Lorsqu'ils sont centrés sur un chercheur, les graphes de relations, permettent de mettre en évidence son environnement collaboratif, que ce soit en interne, accompagnés de non permanents (doctorants, post-doctorants, ingénieurs...), mais également sur la scène locale, régionale, nationale ou internationale. La capacité à s'entourer de collaborateurs sur ces différents périmètres sont très variables d'un chercheur à l'autre et vont bien entendu influencer sur sa production ainsi que sur son rayonnement.

Complémentaire aux graphes de relation, une liste organisée par pays, nommée Places, présente les organismes partenaires des publications.

Bien entendu, ces collaborations sont regardées par le filtre des publications sur HAL. Les représentations graphiques n'intègrent par conséquent pas les collaborations informelles ou naissantes, celles qui n'ont pas nécessairement donné lieu à publication. Des sources complémentaires, ouvertes ou en accès restreint, pourraient compléter ces vues. Des recommandations de collaborations pourraient être également effectuées.

Mesure de la qualité des supports de publication

L'évaluation automatique de la qualité des articles scientifiques d'une collection ou d'une sous-collection est bien entendu un gain de temps précieux mais surtout une source d'information particulièrement intéressante, pour l'évaluation en tant que telle, mais surtout pour les recommandations qui peuvent en découler, en terme de stratégie de publications, de conférences ou de journaux à cibler.

Dans notre prototype, nous avons montré qu'il était possible de se connecter à des sources mesurant la qualité des conférences et des journaux scientifiques, comme CORE et SCImago Journal Rank et Qualis, pour renseigner automatiquement sur la stratégie de publication d'un chercheur, d'une équipe ou d'un laboratoire.

Les premières expériences menées au sein du laboratoire GREYC montrent qu'il est possible de produire de tels indicateurs. Bien entendu, les algorithmes d'appariement sous-jacents mériteraient d'être optimisés et précisément évalués.

Complémentaire, un graphe de type semi-anneau, nommé Audience, permettant d'évaluer la proportion des publications internationales et la proportion des publications nationales, en fonction des médias.

Enfin, une vue détaillée, exportable dans un tableur, permet de superviser les visualisations synthétiques, mais également de produire ses propres représentations graphiques, en corrigeant si nécessaire la qualification automatique.

Analyse des thématiques scientifiques et des domaines d'application ciblés

L'analyse des thématiques scientifiques, croisée avec les domaines d'application retenus ou envisagés, représente un véritable intérêt, en particulier pour les cellules de valorisation ou les industriels en quête d'innovation. Si l'intérêt de produire de telles représentations synthétiques ne fait aucun doute, nous n'avons pour le moment réalisé aucune expérimentation allant dans ce sens. Un état de l'art en la matière sera à effectuer et une collaboration avec des experts de valorisation devrait être menée.

Production liée au projet

Le logiciel fait l'objet d'une étude de valorisation par les services de Normandie Valorisation et n'a par conséquent pas donné lieu à publication.

Le logiciel a fait l'objet de deux dépôt logiciel auprès de l'APP :

- IDDN.FR.001.110014.000.S.C.2020.000.20700 – DATE DU DÉPÔT : 12/03/2020
- IDDN.FR.001.110014.001.S.C.2020.000.20700 – DATE DU DÉPÔT : 25/05/2021

Le site du projet :

<https://aura.normandie-univ.fr/>

GREYC-HULTECH

Last update: January 20 2021 22:07:47 - 312 references.

Media Type

- Conference 149
- Conference/A 22
- Conference/A* 11
- Conference/A2 2
- Conference/B 21
- Conference/B2 2
- Conference/B3 1
- Conference/C 11
- Journal 61
- Journal/Q1 8
- Journal/Q2 9
- Journal/Q3 3
- Journal/Q4 8
- Other 63
- Thesis 16
- Workshop 23
- Workshop/A 6
- Workshop/A* 2
- Workshop/A2 1
- Workshop/B1 1
- Workshop/B5 2
- Workshop/C 1

Year

- 2010 37
- 2011 36
- 2012 56
- 2013 37
- 2014 37
- 2015 25
- 2016 20
- 2017 19
- 2018 15
- 2019 21
- 2020 8
- 2021 1

Document Type

- Article 62
- Communication 166
- Chapitre 20
- Dir. Ouvrage 14
- HR 4
- Autre 10
- Ouvrage 9

Quality Assessment (beta)

- Core: A (74), A* (28), B (13), C (21)
- Qualis: A2 (9), B1 (3), B2 (2), B3 (1), B5 (2)
- Scimago: Q1 (8), Q2 (9), Q3 (3), Q4 (8)

Audience Ranking

- International: 179
- National: 54
- Undefined: 0

Places

Autriche (1)

- Farm Animals and Veterinary Public Health

Allemagne (2)

- LIS
- MPI

Finlande (1)

- Department of Computer Science [Helsinki]

France (64)

- LIA
- CERSA
- LIG [2016-2019]
- Lattice
- IRISA
- Orange Labs
- Orange Labs [Issy les Moulineaux]
- LIUM
- Inria Saclay - Ile de France
- CECOUI
- LLACAN
- IRICE
- Inria Nancy - Grand Est
- ISCC

S	Q	Media	Audience	Year	Venue	Authors	Title	hal
		Journal	International	2021	Proceedings of the Second Workshop on Financial Technology and Natural Language Processing	Emmanuel Giguët, Gaël Lejeune	Daniel at the FinSBD-2 Task: Extracting List and Sentence Boundaries from PDF Documents, a model-driven approach to PDF document analysis	hal
A		Workshop	International	2020	1st Joint Workshop on Financial Narrative Processing and MultiLing Financial Summarisation @COLING 2020	Emmanuel Giguët, Gaël Lejeune, Jean-Baptiste Tanguy	Daniel@FinTOC2 Shared Task: Title Detection and Structure Extraction	hal
		Conference	International	2020	Proceedings of the 12th International ACM Web Science Conference (WebSci '20), Southampton, UK, 17-21 July 2020	Govind, Céline Alec, Marc Spaniol, Amit Kumar	Blogger or President? Exploitation of Patterns in Entity Type Graphs for Representative Entity Type Classification	hal

GREYC-HULTECH

Last update: January 20 2021 22:07:47 - 312 references.

Media Type

- Conference 149
- Conference/A 22
- Conference/A* 11
- Conference/A2 2
- Conference/B 21
- Conference/B2 2
- Conference/B3 1
- Conference/C 11
- Journal 61
- Journal/Q1 8
- Journal/Q2 9
- Journal/Q3 3
- Journal/Q4 8
- Other 63
- Thesis 16
- Workshop 23
- Workshop/A 6
- Workshop/A* 2
- Workshop/A2 1
- Workshop/B1 1
- Workshop/B5 2
- Workshop/C 1

Year

- 2010 37
- 2011 36
- 2012 56
- 2013 37
- 2014 37
- 2015 25
- 2016 20
- 2017 19
- 2018 15
- 2019 21
- 2020 8
- 2021 1

Document Type

- Article 62
- Communication 166
- Chapitre 20
- Dir. Ouvrage 14
- HR 4
- Autre 10
- Ouvrage 9

Quality Assessment (beta)

- Core: A (74), A* (28), B (13), C (21)
- Qualis: A2 (9), B1 (3), B2 (2), B3 (1), B5 (2)
- Scimago: Q1 (8), Q2 (9), Q3 (3), Q4 (8)

Audience Ranking

- International: 179
- National: 54
- Undefined: 0

Places

Autriche (1)

- Farm Animals and Veterinary Public Health

Allemagne (2)

- LIS
- MPI

Finlande (1)

- Department of Computer Science [Helsinki]

France (64)

- LIA
- CERSA
- LIG [2016-2019]
- Lattice
- IRISA
- Orange Labs
- Orange Labs [Issy les Moulineaux]
- LIUM
- Inria Saclay - Ile de France
- CECOUI
- LLACAN
- IRICE
- Inria Nancy - Grand Est
- ISCC

S	Q	Media	Audience	Year	Venue	Authors	Title	hal
		Journal	International	2021	Proceedings of the Second Workshop on Financial Technology and Natural Language Processing	Emmanuel Giguët, Gaël Lejeune	Daniel at the FinSBD-2 Task: Extracting List and Sentence Boundaries from PDF Documents, a model-driven approach to PDF document analysis	hal
A		Workshop	International	2020	1st Joint Workshop on Financial Narrative Processing and MultiLing Financial Summarisation @COLING 2020	Emmanuel Giguët, Gaël Lejeune, Jean-Baptiste Tanguy	Daniel@FinTOC2 Shared Task: Title Detection and Structure Extraction	hal
		Conference	International	2020	Proceedings of the 12th International ACM Web Science Conference (WebSci '20), Southampton, UK, 17-21 July 2020	Govind, Céline Alec, Marc Spaniol, Amit Kumar	Blogger or President? Exploitation of Patterns in Entity Type Graphs for Representative Entity Type Classification	hal

GREYC-HULTECH

Last update: January 20 2021 22:07:47 - 312 references.

Media Type

- Conference 42
- Conference/A 12
- Conference/A* 6
- Conference/A2 1
- Conference/B 8
- Conference/B2 0
- Conference/B3 0
- Conference/C 4
- Journal 10
- Journal/Q1 3
- Journal/Q2 3
- Journal/Q3 0
- Journal/Q4 0
- Other 3
- Thesis 0
- Workshop 10
- Workshop/A 3
- Workshop/A* 1
- Workshop/A2 1
- Workshop/B1 0
- Workshop/B5 2
- Workshop/C 0

Year

- 2010 1
- 2011 6
- 2012 9
- 2013 4
- 2014 14
- 2015 7
- 2016 8
- 2017 3
- 2018 1
- 2019 7
- 2020 5
- 2021 0

Quality Assessment (beta)

- Core 34
- A 15
- A* 7
- B 8
- C 4
- A2 4
- B5 2
- Q1 6
- Q2 3

Audience Ranking

- International 60
- National 2
- Undefined 0

International: 4, 4, 2, 2, 3, 3

National: 3, 3

Places

Search for places...

- France (14)
- LIA
- LIG [2016-2019]
- GREYC
- LIFO
- IRISE
- Hubert Curien
- SITR
- ALLIANSTIC
- LIFO
- IRIT
- LPCN
- STH
- CRI
- LPN
- Irlande (1)
- NCLT
- Portugal (5)
- HULTIG
- CRACS & Inesc TEC [Porto]
- INESC
- LAAD@INESC

Document Type	S	C	Q	Media	Audience	Year	Venue	Authors	Title	hal
Article	10			Journal	International	2020	Micromachines	Fabrice Maurel, Gaël Dias, Waseem Safi, Jean-Marc Rouffon, Pierre Beust	Layout Transposition for Non-Visual Navigation of Web Pages by Tactile Feedback on Mobile Devices	
Communication	51			Other	-	2020		Syed Arbaaz Qureshi, Mohammed Hasanuzzaman, Sripama Saha, Gaël Dias	The Verbal and Non Verbal Signals of Depression – Combining Acoustics, Text and Visuals for Estimating Depression Level	
Chapitre	0			Conference	International	2020	42nd European Conference on Information Retrieval (ECIR)	Nesrine Bannour, Gaël Dias, Youssef Chahir, Houssam Akhmouch	Patch-based Identification of Lexical Semantic Relations	
Dir. Ouvrage	2			Workshop	International	2020	6th Workshop on Semantic Deep Learning (SemDeep-6) associated to 29th International Joint Conference on Artificial Intelligence and 17th Pacific Rim	Jose G. Moreno, Elvys Linhares Pontes, Gaël Dias	CTLR@WIC-TSV: Target Sense Verification using Marked Inputs and Pre-trained Models	

GREYC-HULTECH

Last update: January 20 2021 22:07:47 - 312 references.

Media Type

- Conference 42
- Conference/A 12
- Conference/A* 6
- Conference/A2 1
- Conference/B 8
- Conference/B2 0
- Conference/B3 0
- Conference/C 4
- Journal 10
- Journal/Q1 3
- Journal/Q2 3
- Journal/Q3 0
- Journal/Q4 0
- Other 3
- Thesis 0
- Workshop 10
- Workshop/A 3
- Workshop/A* 1
- Workshop/A2 1
- Workshop/B1 0
- Workshop/B5 2
- Workshop/C 0

Year

- 2010 1
- 2011 6
- 2012 9
- 2013 4
- 2014 14
- 2015 7
- 2016 8
- 2017 3
- 2018 1
- 2019 7
- 2020 5
- 2021 0

Quality Assessment (beta)

- Core 34
- A 15
- A* 7
- B 8
- C 4
- A2 4
- B5 2
- Q1 6
- Q2 3

Audience Ranking

- International 60
- National 2
- Undefined 0

International: 4, 4, 2, 2, 3, 3

National: 3, 3

Places

Search for places...

- France (14)
- LIA
- LIG [2016-2019]
- GREYC
- LIFO
- IRISE
- Hubert Curien
- SITR
- ALLIANSTIC
- LIFO
- IRIT
- LPCN
- STH
- CRI
- LPN
- Irlande (1)
- NCLT
- Portugal (5)
- HULTIG
- CRACS & Inesc TEC [Porto]
- INESC
- LAAD@INESC

Document Type	S	C	Q	Media	Audience	Year	Venue	Authors	Title	hal
Article	10			Journal	International	2020	Micromachines	Fabrice Maurel, Gaël Dias, Waseem Safi, Jean-Marc Rouffon, Pierre Beust	Layout Transposition for Non-Visual Navigation of Web Pages by Tactile Feedback on Mobile Devices	
Communication	51			Other	-	2020		Syed Arbaaz Qureshi, Mohammed Hasanuzzaman, Sripama Saha, Gaël Dias	The Verbal and Non Verbal Signals of Depression – Combining Acoustics, Text and Visuals for Estimating Depression Level	
Chapitre	0			Conference	International	2020	42nd European Conference on Information Retrieval (ECIR)	Nesrine Bannour, Gaël Dias, Youssef Chahir, Houssam Akhmouch	Patch-based Identification of Lexical Semantic Relations	
Dir. Ouvrage	2			Workshop	International	2020	6th Workshop on Semantic Deep Learning (SemDeep-6) associated to 29th International Joint Conference on Artificial Intelligence and 17th Pacific Rim	Jose G. Moreno, Elvys Linhares Pontes, Gaël Dias	CTLR@WIC-TSV: Target Sense Verification using Marked Inputs and Pre-trained Models	