

Méthodologie HALathon CentraleSupélec 2020

A destination de tous ceux qui veulent savoir ce qui se cache derrière

Amandine Lustrement

Université Paris-Saclay, Pôle IST-Recherche, CentraleSupélec, 9 rue Joliot-Curie, Gif-sur-Yvette, France

Récupération des données

Pour l'occasion, a été créée une collection cachée "HALATHON-CS", paramétrée comme suit :

- Elle rassemble les collections correspondant aux laboratoires qui participent au HALathon ;
- Elle ne contient pas les thèses, HDR, cours, logiciels, vidéos et documents de type cartes

La requête exacte est : `collection_t:(CVN OR EM2C OR FED-MATH OR GEEPS OR LANEAS OR LGI OR LGPM OR LPQM OR MAS OR MICS OR MSSMAT OR SPMS OR SUP_CIDRE OR SUP_IETR OR SUP_IMS OR SUP_LMOPS OR SUP_LSS OR SUP_SONDRA OR UMI-COMPUTERSCIENCE OR UMI-GTL OR SUP_LGEP) AND NOT docType_s:(THESE OR HDR OR LECTURE OR SOFTWARE OR VIDEO OR MAP)`

Les données sont extraites avec OcdHAL, outil permettant d'interroger les collections HAL sous forme de tableur.

Corrections :

- Le CVN ne participant pas au HALathon, celui-ci n'aurait pas dû être dans la requête. Il a été retiré de la liste des publications après extraction de la liste, en supprimant le tampon seul CVN dans Excel ;
- L'historique du GeePs, SUP_LGEP, avait été rajouté pour pouvoir calculer la coupe OpenAccess mais le tampon n'était pas paramétré dans OcdHAL dans la collection HALathon-CS. Les notices étant bien dedans, elles ont été récupérées à partir de l'affiliation directement dans Excel aussi

Le fichier HALATHON-CS186.csv est celui à partir duquel les informations ont été extraites. Pour la moyenne de l'école, c'est le fichier HALATHON-CS186-sans-cvn.csv qui a été utilisé.

Pour chaque coupe et les bonus, les notices ont été filtrées suivant les valeurs indiquées ci-dessous, puis comptées :

Pour la coupe HALathon :

	Nombre de notices	Nombre de documents déposés pendant le HALathon	Nombre de documents total	Nombre de notices avec lien OA
Années	2015-2019	2015-2019	2015-2019	2015-2019
Type de dépôt	Notice	Fichier + Annexe	Fichier + Annexe	Notice
Type de document	ART et COMM	ART et COMM	ART et COMM	ART et COMM
Actes (COMM)	Oui + Vides	/	/	Oui + Vides

Soumis le	/	Entre le	/	/
		27/02/2020 et		
		30/09/2020		

Pour la coupe OpenAccess :

Nombre de notice : champs OCdHAL	
Années	/
Type de dépôt	Notice
Type de document	ART et COMM
Actes (COMM)	Oui + Vides
Soumis le	/
Nombre de notice OA : champs OCdHAL	
Années	/
Type de dépôt	Notice
Type de document	ART et COMM
Actes (COMM)	Oui + Vides
Soumis le	/
linkExtId	openaccess, arxiv, pubmedcentral
Nombre de documents: champs OCdHAL	
Années	/
Type de dépôt	Fichier + Annexe
Type de document	ART et COMM
Actes (COMM)	/
Soumis le	/

Pour les bonus (hors coupe) :

BONUS	Articles et communications	Autres publications	Chapitres d'ouvrages	Directions d'ouvrages	Ouvrages	Posters	Pré-publications	Rapports
Années	/	/	/	/	/	/	/	/
Type de dépôt	Fichier + Annexe	Fichier + Annexe	Fichier + Annexe	Fichier + Annexe	Fichier + Annexe	Fichier + Annexe	Fichier + Annexe	Fichier + Annexe
Type de document	ART+COMM	Autre publi	Chapitre d'ouvrage	Direction d'ouvrage	Ouvrages	Posters	Pré-publications	Rapports
Soumis le	entre le 17/02/20 et 30/09/2020	entre le 17/02/20 et 30/09/2020	entre le 17/02/20 et 30/09/2020	entre le 17/02/20 et 30/09/2020	entre le 17/02/20 et 30/09/2020	entre le 17/02/20 et 30/09/2020	entre le 17/02/20 et 30/09/2020	entre le 17/02/20 et 30/09/2020

L'ensemble de ces informations est aussi indiqué dans Ratios finaux v2.xlsx.

Calcul des scores

Coupe HALathon

1. TI intégral sur la période 2015-2019

Afin que soient pris en compte les efforts globaux sur la période, de ne pas léser les laboratoires qui avaient peu de fichiers restants à déposer, et pour avantager les laboratoires qui ont atteint (ou sont proches de) 70% -l'objectif initial de l'événement- est pris en compte la part de texte intégral disponible au regard de l'ensemble des notices articles et communications dans HAL sur la période 2015-2019. De plus, potentiellement, les laboratoires ayant déjà beaucoup déposé de fichiers dans HAL peuvent se retrouver avec beaucoup de fichiers difficiles à retrouver, voire uniquement des documents qui n'ont pas pu être déposés auparavant pour cette raison. Leur effort pendant le HALathon (2.) sera donc forcément moindre, peu importe la volonté qu'on peut y mettre.

2. Part de TI déposée sur la période du HALathon

Il s'agit de la part de fichiers déposés à partir de la date de début du HALathon, 17 février 2020, qui correspond au premier envoi de mail pour le premier laboratoire (la campagne de communication ayant commencé bien en amont). Le pourcentage est calculé sur la part de fichiers demandés pour chaque laboratoire, à savoir les notices d'articles et communication avec actes, parus entre 2015 et 2019. Ce point montre l'effort fourni pendant le HALathon pour les laboratoires, même si tout ne se joue pas là : ainsi, nous savons que des chercheurs n'ont plus les fichiers acceptés pour publication, ont fait l'effort de les rechercher auprès de leurs anciens co-auteurs sans succès ; le chiffre peut aussi ne pas refléter les efforts fournis quand des chercheurs "gros publiants" ont quitté le laboratoire et n'ont pas transmis de fichiers, même quand ils ont été sollicités.

3. Participation

Il s'agit de compter la part de participants au HALathon au regard du nombre de "publiants" contactés pour chaque laboratoire.

Sont comptabilisés comme participants : les personnes venues au stand HALathon organisé dans leur laboratoire, les RDV individuels, les RDV Teams, les contacts par mail avec envoi de fichiers ou qui ont autorisé la recherche et dépôt en leur nom.

Sont comptabilisés comme publiants : Les chercheurs, enseignants-chercheurs, maîtres de conférences, ingénieurs, post-doctorants et doctorants qui travaillent encore à CentraleSupélec, et qui ont été contactés dans ce cadre, soit par nos mails individuels, soit par leur directeur de laboratoire.

Ne sont pas comptabilisé comme publiants du laboratoire ceux qui n'y travaillent plus, même si parfois beaucoup de publications leur appartiennent et même s'ils ont été contactés.

Les chercheurs qui ont quitté le laboratoire mais ont quand même participé sont comptabilisé à la fois dans le nombre de publiants et dans le nombre de participants.

La participation ne peut pas être prise comme critère seule, n'a pas non plus été réhaussée par rapport aux deux autres (voir ci-dessous), pour plusieurs raisons : un seul et même chercheur pouvait dans certains labos remplir un tiers des publications demandées, VS plusieurs chercheurs pouvaient, en mettant chacun 1 ou 2 document(s) atteindre le même objectif. Avec la COVID-19, nous sommes également conscientes qu'une grande partie des publiants ont eu à transposer leurs enseignements à distance (grosse charge de travail supplémentaire), en plus de s'adapter à la situation pour le reste de leurs travaux ou tout simplement de veiller sur leur famille.

Chacun de ces critères amène un pourcentage, qui est traduit en “points” allant de 0 à 10 (55% = 5,5). Les trois critères ajoutés donnent un score sur 30. Les laboratoires ayant le score le plus élevé sont ceux qui remportent les places du podium.

Coupe OpenAccess

La coupe OpenAccess vient récompenser le laboratoire qui s’investit le plus dans l’accès ouvert à ses publications. L’ensemble des années, depuis la création du laboratoire est pris en compte. Les types de documents retenus restent les articles et les communications.

Un seul critère est retenu : la part de publications disponible sur HAL + la part de publications disponibles en OpenAccess (ArXiv, PubMed et voie dorée), au regard de l’ensemble de la production. Celui qui détient le pourcentage le plus élevé remporte la coupe.

Bien que les laboratoires n’aient pas été créés en même temps, le nombre de notices antérieures à l’année 2000 est relativement faible. A côté de cela, des fichiers ont été déposés pour certaines des publications antérieures à 2000. Par conséquent, aucun ajustement n’a été fait.