

Guide à l'utilisation du logiciel Publish or Perish

Introduction au logiciel Publish or Perish

Objectifs du logiciel

« **Publish or Perish** » (PoP) est un logiciel qui recherche et analyse des citations académiques à partir de **Google Scholar**. A partir de cet ensemble de citations, il calcule un certain nombre de statistiques (nombre d'articles, de citations, h-index, g-index, etc.) qui vont permettre d'évaluer le chercheur qui en est l'auteur.

Ce logiciel a été conçu afin de permettre aux chercheurs de présenter sous son meilleur jour l'influence qu'ils ont eue sur leur domaine scientifique.

Les chiffres obtenus n'ont donc pas pour ambition de permettre la comparaison pure et simple d'un ensemble de chercheurs entre eux. En effet, la signification de ces chiffres peut varier en fonction de nombreuses circonstances (voir les « précautions d'interprétation » en page 5 de ce guide).

Associé à d'autres outils similaires (ceux du Web of Science ou de Scopus par exemple), Publish or Perish va néanmoins permettre de **se faire une idée de l'impact du chercheur** au sein de la communauté scientifique.

Installer Publish or Perish

Pour **Windows**, téléchargez le logiciel à cette adresse :

<http://www.harzing.com/download/PoPSetup.exe> .

Cette version est compatible avec Windows 2000, XP, 2003, Vista et 7.

Ouvrez le fichier téléchargé et suivez les instructions de l'installeur automatique.

Pour **Linux**, téléchargez le logiciel à cette adresse :

<http://www.harzing.com/download/pop-linux.tgz>

Cette version est compatible avec des versions récentes de Linux (elle a été testée sur Ubuntu 8.04). Elle est basée sur la bibliothèque GTK+2.

Décompressez le fichier et installez le.

Utilisation du logiciel

Utiliser Publish or Perish

Pour obtenir les statistiques d'un auteur, commencez par cliquer sur l'onglet « **Author impact analysis** », dans la partie gauche de la page.

Puis entrez le nom de l'auteur dans le champ « **Author's name** » (sous la forme Nom, Initiale du prénom) et cliquez sur « Look Up ».

Au bout de quelques secondes, la liste des publications trouvées s'affiche dans la partie basse de la page et les résultats calculés se retrouvent sous le terme « **Results** ».

Passez en revue les publications trouvées. Il y a de fortes chances pour que nombre d'entre elles ne correspondent pas réellement au chercheur souhaité.

Afin d'**affiner** votre recherche, vous pouvez :

- Mettre le nom et l'initiale du chercheur entre **guillemets**. Remarque : « Nom, P » donnera le même résultat que « Nom P », « P Nom » ou « P, Nom ».
- **Excluez** certains noms en les entrant dans le champ « Exclude these names ».
- Affinez à **certaines années** en remplissant les champs « Years of publication between » ... « and » ...
- Affinez à **certain domaines** parmi ceux proposés en haut à droite de la fenêtre. Seuls les domaines cochés seront pris en compte.
- Enfin, vous pouvez choisir d'**enlever à la main** certaines publications de la liste obtenue en décochant la case « Cites » en début de ligne.

Remarque : pour cocher toutes les publications, utilisez le bouton « Check all » ; pour les décocher, utilisez le bouton « Uncheck all ». Vous pouvez aussi sélectionner plusieurs publications et les cocher ou décocher d'un seul coup en utilisant les boutons « Check selection » et « Uncheck selection ».

Remarque : les statistiques résultantes sont automatiquement recalculées par le logiciel au fur et à mesure que la liste de publications est affinée.

Vous pouvez exporter les données obtenues dans le menu « File », en cliquant sur l'un des quatre boutons « Save as ... ».

Les quatre formats d'export possibles sont BibTeX (un format utilisé pour les références bibliographiques), CSV (format texte où les données sont séparées par des virgules), EndNote Import (pour le logiciel EndNote) et RIS (un format utilisé notamment par le logiciel Reference Manager).

Le bouton « Copy statistics » permet de copier les statistiques obtenues dans le presse-papier de Windows (au format CSV).

Remarque : il est possible de faire une recherche pour plusieurs auteurs à la fois.

Aller plus loin

Comprendre les résultats obtenus

Voici le détail des résultats statistiques proposés par Publish or Perish.

Nombre total de publications trouvées (« **Papers** »).

Nombre total de citations trouvées (« **Citations** »).

Nombre d'années sur lesquelles s'étalent ces publications (« **Years** »).

Nombre moyen de citations par an (« **Cites/year** »).

Nombre moyen de citations par publication (« **Cites/paper** »).

Nombre moyen de publications par auteur (« **Papers/author** »).

Le nombre moyen d'auteurs par publication (« **Authors/paper** ») permet de se faire une idée sur l'influence de co-auteurs dans le travail du chercheur.

Le **h-index**, proposé par J.E. Hirsh en 2005¹, cherche à mesurer l'impact cumulatif des publications d'un chercheur en se basant sur les citations que son travail a reçues. Un scientifique possède un h-index de la valeur h si h de ses publications ont été citées au moins h fois.

Il est ici complété par la constante **a** (où $N = ah^2$ avec N nombre total de publications du chercheur) et le paramètre **m** (où $h \sim mn$ avec n représentant le nombre d'années depuis la première publication).

Le **g-index**, proposé par Leo Egghe en 2006², est défini comme suit :

Pour un ensemble d'articles classés par ordre décroissant de nombre de citations qu'ils ont reçues, le g-index est le nombre le plus grand pour lequel les g premiers articles ont reçu ensemble au moins g^2 citations.

Cet indice cherche à améliorer le h-index en donnant plus de poids aux articles les plus cités.

Le **h-index contemporain**, proposé en 2006³, ajoute une pondération de chaque article en fonction de son âge, donnant ainsi moins de poids aux articles les plus anciens. Ici, les citations d'un article paru cette année comptent 4 fois plus qu'un article publié il y a quatre ans.

Cet indice est noté « **hc-index** » ou « contemporary h-index ».

¹ Voir <http://arxiv.org/abs/physics/0508025>.

² Leo Egghe. **Theory and practice of the g-index**, *Scientometrics*, Vol. 69, No 1 (2006), pp. 131-152

³ Voir <http://arxiv.org/abs/cs.DL/0607066>

Le **h-index individuel**, proposé lui aussi en 2006⁴, cherche à réduire les effets dus aux co-auteurs. Il est noté « **hl-index** ».

PoP propose aussi un h-index alternatif appelé « **hl,norm** ». Plutôt que de diviser le h-index total par le nombre d'auteurs (comme dans le hl-index), cet indice normalise le nombre de citations pour chaque publication en divisant le nombre de citations par le nombre d'auteurs du papier. Le hl,norm est ensuite calculé comme le h-index. Cette approche permet d'obtenir un résultat plus fin tout en prenant en compte l'existence des co-auteurs.

AWCR signifie « Age-weighted citation rate ». Inspiré du travail de Bihui Jin⁵, il mesure le nombre de citations d'un ensemble de publications, ajusté en fonction de l'âge de chaque publication individuelle. Le nombre de citations d'une publication est alors divisé par son âge.

Le **AW-index** correspond à la racine carrée du AWCR, ce qui permet sa comparaison avec le h-index. Il vaut environ le h-index si le taux moyen de citation reste plus ou moins constant avec les années.

Le **AWCRpA** (per-author age-weighted citation rate) correspond à l'AWCR mais est normalisé en fonction du nombre d'auteurs de chaque publication.

Le **e-index**, proposé par Chun-Ting Zhang en 2009⁶, cherche à différencier les scientifiques ayant un même h-index mais ayant différents modèles de citation.

Le « **hm-index** » est une variation du hl-index, proposée par Michael Schreiber en 2008⁷. Cet indice utilise un comptage fractionnel des publications plutôt que de réduire le compte de citations, déterminant ainsi un index multi-auteur.

Les résultats se terminent par une courte analyse du nombre d'auteurs par publication (nombre de publications avec X auteurs).

Précautions quant à l'interprétation des résultats

Il est nécessaire de rester prudent quant à l'analyse de ces résultats. Si un chercheur obtient de bonnes statistiques, alors il a très probablement eu un impact significatif dans son domaine. Mais l'inverse est loin d'être vrai. Si ses statistiques sont faibles, alors peut-être travaille-t-il dans un domaine confidentiel ou publie-t-il dans des langues autres que l'anglais, ce qui réduit le champ de publications et de citations récupérées par Google Scholar.

⁴ Pablo D. Batista, Monica G. Campiteli, Osame Kinouchi, and Alexandre S. Martinez. **Is it possible to compare researchers with different scientific interests?**, *Scientometrics*, Vol 68, No. 1 (2006), pp. 179-189

⁵ Bihui Jin. **The AR-index: complementing the h-index**, *ISSI Newsletter*, 2007, 3(1), p. 6

⁶ Chun-Ting Zhang. **The e-index, complementing the h-index for excess citations**, *PLoS ONE*, Vol 5, Issue 5 (May 2009)

⁷ Michael Schreiber. **To share the fame in a fair way, h_m modifies h for multi-authored manuscripts**, *New Journal of Physics*, Vol 10 (2008), 040201-1-8

De plus, l'utilisation de Google Scholar tend à sous-estimer les citations en sciences sociales et sciences humaines particulièrement, car il ne prend pas toujours en compte les publications d'ouvrages ou de chapitres d'ouvrages. Néanmoins, Google Scholar couvrant un plus large panel de journaux dans ces domaines, il permet d'obtenir un plus grand nombre de citations qu'en utilisant le Web of Science.

Par contre, l'ISI couvre particulièrement bien les domaines des sciences de la nature et de la santé. Les nombres de citations obtenues peuvent donc être supérieures à celles trouvées par Publish or Perish.

Au final, il est conseillé d'utiliser et de comparer plusieurs outils (Publish or Perish, mais aussi les outils du Web of Science ou de Scopus) afin de pouvoir profiter de chiffres optimaux.

En savoir plus

Pour en savoir plus sur le logiciel Publish or Perish, vous pouvez consulter le site Web de sa conceptrice, Anne-Wil Harzing : <http://www.harzing.com/pop.htm>.