

Approche théorique des indicateurs de consultation sur le Web

Thierry, Lafouge, ELICO, EA 4147
Université de Lyon, Université Lyon1 Villeurbanne, F-69622, France

1^{er} décembre 2008

Table des matières

1	Les grandeurs mesurables et les unités en infométrie	3
1.1	Les différentes unités	3
1.2	Les unités dérivées simples	4
1.3	Les unités dérivées, les taux	4
2	Ressources électroniques dans les bibliothèques : grandeurs mesurables et unités	5
3	Le référentiel de l' ISI, Web of knowledge	6
3.1	Impact factor	6
4	Construction du « Journal usage impact factor »	7
4.1	Construction d'une fonction pour un type d'item	7
5	Distributions d'usage, lois puissances inverses	8
5.1	Exemple de l'usage des revues "majeures" de l'Inist	8
5.2	Les distributions statistiques des commandes	9
5.3	Identifier le coeur de l'information à travers les usages	9

1 Les grandeurs mesurables et les unités en infométrie

1. Le signal *Bit* *S*
2. Le temps *Seconde* *t*
3. Le texte *Mot* *T*
4. La population *Individu* *G*
5. L'utilité *Usage* *U*

1.1 Les différentes unités

[Lafouge et al., 2002]

- Mot = forme linguistique minimale indépendante « douée de sens »
- Individu = auteur, lecteur, visiteur,...
- Usage = Nb d'emprunts, de consultations, de citations,....
d'un « objet informationnel »

1.2 Les unités dérivées simples

- Le nombre d’usages par individu $\frac{U}{G}$
- Le nombre de mots *consommés* par individu $\frac{T}{G}$
- le nombre d’usages par unité de temps $\frac{U}{t}$
fréquence d’usage de l’information **Usagivité**

1.3 Les unités dérivées, les taux

- Taux de croissance. Soit $t_1, t_2, t_2 > t_1$ deux dates définissant une période
On définit le taux de croissance :

$$\frac{U(t_2) - U(t_1)}{U(t_1)}$$

- 5000 consultations en 2003 à 15000 consultations en 2008
Le taux de croissance est de :

$$\frac{15000 - 5000}{5000} = 2$$

2 Ressources électroniques dans les bibliothèques : grandeurs mesurables et unités

- item= « Partie identifiable de manière unique d'une oeuvre publiée qui peut être »
 1. une oeuvre originale
 2. un article
 3. un résumé

Les formats PDF, HTML....., d'un même article (par exemple) seront comptabilisés comme des items différents. (COUNTER) [Charbonnier, 2007]

- individu = auteur, lecteur, visiteur : **identification N° IP** :
Problème de granularité : poste, le laboratoire, la discipline.....
Nécessité d'une **politique d'identification** via les N°IP.
- usage = nombre de téléchargements, de consultations...
 - Revues
 - Bases de données bibliographiques
 - Encyclopédies....

Nécessité d'une politique via un **proxy documentaire**.

3 Le référentiel de l' ISI, Web of knowledge

Les indicateurs du *Journal Citation Reports*

- *Total cities* : nombre total de citations,
- *Impact factor* (voir le recueil d'articles de *Scientometrics* [Braun, 2007]),
- *Immediaty index* : rapidité de l'impact d'un article,
- *Cited Half life* : longévit  d'un l'article.

3.1 Impact factor

Exemple du journal **IPM** en 2007¹

- 2005, **IPM** publie 87 articles,
- 2006, **IPM** publie 101 articles,
- En 2007, 157 citations d'articles de 2005,
- En 2007, 125 citations d'articles de 2006,

Par d finition l'impact factor **CIF**

$$CIF = \frac{157 + 125}{87 + 101} = 1.5$$

¹Donn es extraites du JCR le 23 novembre 2008

4 Construction du « Journal usage impact factor »

On construit des indicateurs d'usage [Rowland and Nicholas, 2007]
On copie l'ISI "Citation=Téléchargement".

Soit **UIF** l'indicateur d'usage identique à l'impact factor **CIF**.

Soit **J** un journal, une date **t** on se donne une fenêtre de, 2 ans
on calcule alors le **UIF** du journal **J**

$$UIF = \frac{U_{t-1}(t) + U_{t-2}(t)}{A(t-1) + A(t-2)}$$

Le numérateur correspond à n'importe quel item télédéchargé,....

Le dénominateur est limité.....

Soit **UII** l'indicateur d'usage semblable à l'*immediaty index*.

$$UII = \frac{U_t(t)}{A(t)}$$

4.1 Construction d'une fonction pour un type d'item

De tels indicateurs nécessitent la mise en place de fonctions.

$$U_{Item(t')}(t)$$

Nombre d'usages à l'instant t pour un type d'item "daté" t'

Il n'existe pas de recommandation COUNTER.

5 Distributions d'usage, lois puissances inverses

Les distributions d'usage sont en général des lois de type puissances inverses.
On les appellent sous des noms diverses suivant le domaine
Pareto, Lotka, Bradford, Zipf [Egghe, 2005]

5.1 Exemple de l'usage des revues "majeures" de l'Inist

ANNÉE	Nb de revues	Nb de commandes	Entropie	loi des 80% 20%
1997	4,876	431,833	0.9	28.9
1998	5,074	443,099	0.9	28.9
1999	5,243	475,142	0.9	29.11
2000	5,390	481,255	0.9	29.24
2001	5,493	458,379	0.9	29.42
2002	5,561	413,136	0.9	29.4
2003	5,673	366,939	1	29.74
2004	5,778	340,248	0.9	29.01
2005	5,951	288,090	1	28.79
2006	5,934	244,888	0.9	28.7

TAB. 1 – Volume des commandes des revues "majeures de l'INIST"

5.2 Les distributions statistiques des commandes

5.3 Identifier le coeur de l'information à travers les usages

Coupure : 1268 revues commandées plus de 97 fois

Soit 26% des revues assurent 77% des commandes.

Utilisation de méthodes "semi-automatique" pour couper la distribution en 3 zones.

Ces méthodes sont en cours de validation sur des distributions de vocabulaire.

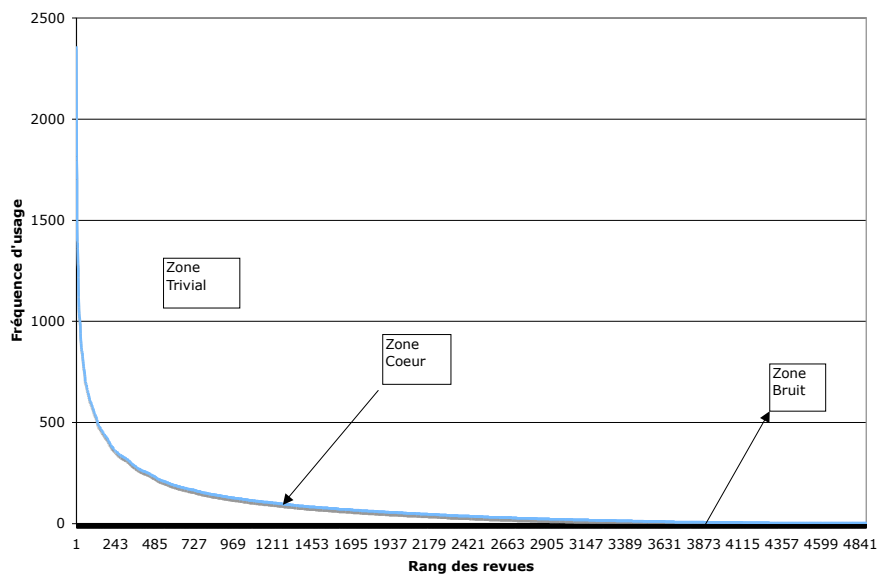


FIG. 1 – Distribution des commandes des revues en 1997

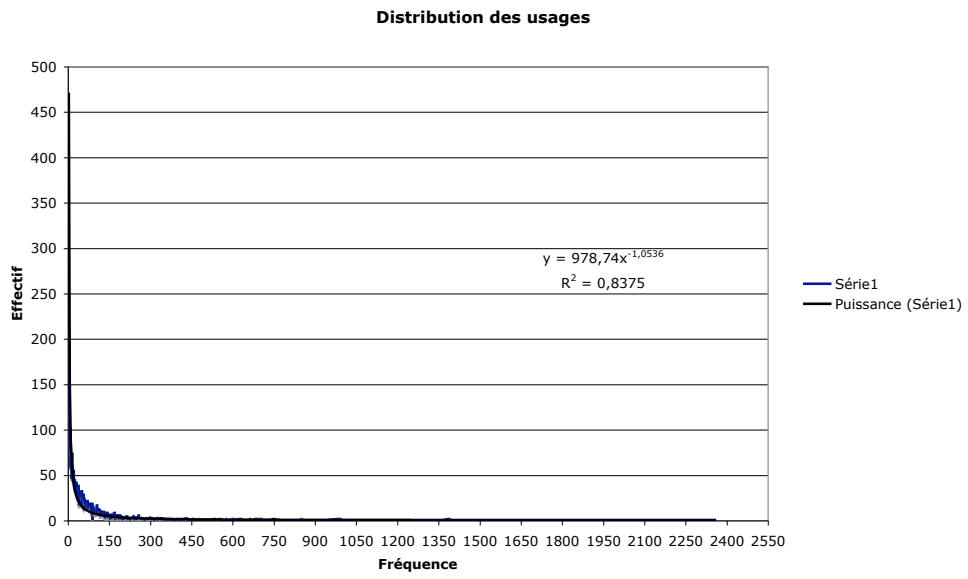


FIG. 2 – Distribution des revues en 1997

Références

- [Braun, 2007] Braun, T. (2007). *The Impact factor of Scientific and Scholarly Journals. Is Use and Misuse in Research Evaluation*, volume 2. AKADÈMIA KIADO.
- [Charbonnier, 2007] Charbonnier, F. (2007). Le recueil de données statistiques sur l’usage des ressources électroniques au service commun de la documentation de l’université claudes bernard lyon1. Technical report, Service Commun de la Documentation de l’université Claude Bernard Lyon1.
- [Egghe, 2005] Egghe, L. (2005). *Power laws in the information production process : lotkaian informetrics*. Elsevier.
- [Lafouge et al., 2002] Lafouge, T., LeCoadic, Y., and Michel, C. (2002). *Éléments de statistique et de mathématique de l’information*. Presses de l’enssib.
- [Rowland and Nicholas, 2007] Rowland, I. and Nicholas, D. (2007). The missing link : journal usage metrics. *Aslib Proceedings : New Information Perspectives*, 59(3) :222–228.