

Algorithme PageRank de Google

L'objectif de cette activité est d'expliquer une version simplifiée de l'algorithme PageRank de Google qui fournit un classement des pages web par ordre d'intérêt (indépendamment de la recherche effectuée).

L'activité est soutenue par des diapositives, un programme JAVA et comprend une expérience effectuée par le public.

Déroulement de l'activité :

1. On débute par une petite discussion générale sur le fonctionnement de Google et la nécessité de pouvoir classer un grand nombre de pages automatiquement.
2. On discute ensuite plus précisément des critères à adopter pour classer les pages en utilisant juste les informations fournies par le réseau (sans jugement humain). On s'appuie sur la page d'information fournie par Google.
3. On modélise ensuite le web par un graphe : les sommets représentent les pages web tandis que les arêtes (orientées) indiquent les liens hypertextes entre les pages. On prendra en exemple un « mini-web » de cinq pages.
4. Avec l'aide des diapositives, on explique l'algorithme : les pages les mieux classées sont celles sur lesquelles on passera le plus souvent si on surfe sur le web en passant d'une page à l'autre en suivant à chaque fois un lien hypertexte au hasard.
5. Le public réalise une simulation sur l'exemple du « mini-web » de cinq pages, le choix du lien à suivre est déterminé par un lancer de dé.
6. On effectue la même simulation grâce au programme JAVA. Cela permet d'effectuer un grand nombre d'itérations sur ce premier exemple. On peut aussi effectuer la simulation en supprimant un lien afin de voir ce que cela change sur le classement. Grâce à d'autres exemples de réseaux, on peut également mettre en valeur quelques petits défauts de cet algorithme (dissimulés dans le premier exemple) et discuter la façon d'y remédier.
7. On termine en répondant aux éventuelles remarques et questions du public.

Remarques concernant l'expérience réalisée par le public :

1. Le graphe est dessiné sur un plateau en carton, un pion représentant le « surfeur ». A chaque changement de page, un lancer de dé détermine lequel parmi les liens sortant de la page en cours sera suivi. Le compte des visites est effectué en laissant à chaque passage un mot ou une phrase en rapport plus ou moins direct avec les mathématiques à la personne représentant le site visité (comme un commentaire laissé sur un blog...)
2. Le public a été divisé en deux groupes et l'expérience est réalisée indépendamment par chaque groupe, pour voir que malgré l'aléa c'est au final le même classement qui se dégage.
3. Il peut être intéressant de faire une parenthèse sur le fait qu'il n'y a pas besoin que chaque page ait 6 liens sortants pour pouvoir faire l'expérience avec un dé à 6 faces. S'il est à peu près évident qu'on peut choisir un lien parmi deux ou trois avec un tel dé, il est bon de remarquer que c'est tout à fait possible avec quatre, cinq ou même plus de six liens sortants...