

pages d'une catégorie sont plus souvent liées entre elles qu'à celles de l'autre communauté.

Plus généralement, *Clever* révèle la structure sous-jacente du réseau. Bien que ce dernier ait crû de façon fulgurante et anarchique, il possède une structure intrinsèque – bien qu'imparfaite – fondée sur les liaisons entre pages.

Le lien avec l'analyse de citations

Par son principe, *Clever* ressemble à l'analyse de citations, l'étude de la manière dont les articles scientifiques se citent : la principale mesure de l'importance des articles est le «facteur d'impact», qui juge les publications d'après le nombre de références que les autres articles en font. Mis au point par Eugene Garfield, un spécialiste de l'information scientifique, le facteur d'impact préside à l'«Index des citations scientifiques» (*Science Citation Index*).

Sur le réseau, le facteur d'impact d'une page pourrait être le nombre de liens qui pointent vers elle. Toutefois, cette approche simpliste favoriserait, même pour des requêtes sur des sujets spécifiques, les pages généralistes très connues, telle la page d'accueil du *Monde*.

Dans le domaine de l'analyse des citations, les chercheurs ont tenté d'améliorer la mesure de E. Garfield, qui compte de la même façon toutes les références. Ne serait-on pas avisé de donner plus de poids aux citations d'un journal plus important ? Le problème, avec cette approche, est qu'on donne une définition circulaire du mot «importance», comme dans la définition des pivots et des références.

Dès 1976, Gabriel Pinsky et Francis Narin, du Laboratoire de recherches du CHI, dans le New Jersey, ont éliminé cet obstacle en mettant au point une méthode itérative de calcul qui aboutit à un ensemble stable de notes, qu'ils nomment poids d'influence. Toutefois, ils ne distinguent pas référence et pivot.

Cette différence révèle une particularité entre les pages Internet et la littérature scientifique classique : dans le cyberspace, des sites de référence en compétition, comme *Netscape* et *Microsoft*, ne reconnaissent pas leur existence mutuelle et ne peuvent donc être reliés que par un ensemble de pivots tiers ; inversement, les revues scientifiques rivaux pratiquent des

citations croisées qui rendent le rôle des pivots moins crucial.

D'autres équipes utilisent différemment les hyperliens, pour la recherche sur le réseau. À l'Université de Stanford, Serge Brin et Larry Page ont mis au point un moteur de recherche, nommé *Google*, qui utilise un classement fondé sur les poids d'influence de G. Pinsky et F. Narin. L'équipe de Stanford considère un internaute qui suit des liens en faisant un certain nombre de sauts au hasard, qui le mènent à certaines pages plus fréquemment qu'à d'autres. Ainsi *Google* trouve un seul type de pages universellement importantes, intuitivement celles qui sont fréquemment visitées lors d'une marche au hasard dans la structure des liens du réseau. Concrètement, pour chaque page, *Google* additionne les notes des sites qui pointent vers la page. C'est pourquoi, lors d'une requête spécifique, *Google* répond rapidement en trouvant toutes les pages qui contiennent le texte recherché et en les ordonnant selon leur rang prédéterminé.

Les moteurs *Google* et *Clever* diffèrent cependant par deux aspects importants. D'une part, *Google* élabore un ordre de classement initial et l'utilise quelle que soit la requête, alors que *Clever* construit un ensemble de base différent pour chaque terme recherché, puis fixe les priorités des pages dans le contexte de cette requête. Au total, *Google* répond plus rapidement. D'autre part, *Google* ne regarde que dans une seule direction, d'un lien vers un autre, alors que *Clever* regarde aussi dans l'autre direction, pour savoir quels sont les sites qui pointent vers une page de référence. *Clever* tire ainsi parti du phénomène sociologique qui fait que les humains construisent naturellement des contenus qui expriment leur expertise sur un sujet donné.

La recherche continue

Nous cherchons encore à améliorer *Clever*. Dans notre démarche, nous cherchons à associer texte et hyperliens. Pour ce faire, nous estimons que certains liens ont plus de poids que d'autres suivant la pertinence du texte qui se trouve en regard. Plus précisément, nous analysons le contenu des pages dans l'ensemble de base et nous cherchons le nombre d'occurrences et les positions relatives du terme recherché. Ces informations pondèrent ensuite certaines connexions entre

pages. Si, par exemple, le texte de la requête apparaît fréquemment à proximité d'un lien, alors la pondération correspondante est augmentée.

Nos premières expériences indiquent que ce raffinement améliore substantiellement la pertinence des résultats (une imperfection de *Clever* a parfois l'inconvénient d'élargir trop sa recherche, quand la requête est très précise : ainsi, si l'on cherche la Maison Ozenfant, de Le Corbusier, le système élargit sa recherche et retrouve des informations sur un sujet général comme l'architecture contemporaine). Nous expérimentons d'autres améliorations qui tiennent compte des divers styles de réalisation de pages Internet.

Nous avons aussi commencé à construire automatiquement des listes de ressources Internet, similaires aux annuaires de sites élaborés manuellement par les employés de sociétés comme *Yahoo!* ou *Infoseek*. Les premiers résultats sont aussi bons que ceux des annuaires. De surcroît, ce travail nous a montré que le réseau est plein de petites communautés très liées, avec des intérêts particuliers (les amateurs de sumo, les passionnés de yo-yo de la Creuse ou les amateurs de sauce ketchup au petit déjeuner). Nous recherchons comment découvrir automatiquement ces communautés cachées (voir la figure 1).

Le réseau Internet est aujourd'hui très différent de ce qu'il était il y a seulement cinq ans. Personne ne peut dire ce qu'il sera dans cinq ans. L'indexation du réseau deviendra-t-elle impossible ? Et comment la notion de recherche changera-t-elle alors ? Aujourd'hui, la seule chose dont nous soyons certains est que la croissance du réseau continuera à lancer des défis informatiques à ceux qui veulent naviguer dans cet océan d'informations.

Le projet *Clever* : Soumen Chakrabarti, Byron Dom, Ravi Kumar, Prabhakar Raghavan, Sridhar Rajgopalan et Andrew Tomkins travaillent au Centre de recherche Almaden d'IBM, à San Jose (Californie). Jon Kleinberg travaille au département d'informatique de l'Université Cornell. David Gibson termine sa thèse au département d'informatique de l'Université de Berkeley.

De nombreux liens pour cet article sont disponibles sur notre site Web : <http://www.pourlascience.com/numeros/pls-262/clever.htm>



Le Quercy tropical

C. MOURER-CHAUVIRÉ • S. LEGENDRE • B. MARANDAT • B. SIGÉ

Au début de l'ère tertiaire, la faune du Sud-Ouest de la France ressemblait à celle des régions tropicales actuelles. Les fossiles révèlent l'histoire du climat sur 35 millions d'années.



1. CE PAYSAGE TROPICAL aurait pu être observé près du site d'Itardiès, dans le Tarn-et-Garonne, il y a 30 millions d'années. Le Soleil se couche. Nous sommes placés sous un porche rocheux, à l'entrée d'une grotte creusée dans le calcaire. Sur le plateau, le sol est partiellement couvert d'une argile rougeâtre formée par la dissolution et l'altération chimique de la roche calcaire. La végétation clairsemée, herbacée en touffes et faiblement arborée correspond à un climat assez chaud et aride. Les animaux qui peuplent ce paysage sont très différents de ceux que l'on trouve aujourd'hui dans la région, en particulier les oiseaux, dont les formes apparentées actuelles vivent surtout sous les tropiques. Au premier plan figurent, de gauche à droite : (1) *Issiodoromys medius*, rongeur de la famille des théridomyidés, représentant un grand groupe endémique et éteint du Paléogène d'Europe (il y a 65 à 23 millions d'années) ; (2) deux chauves-souris à feuillets nasaux, d'un genre éteint de la famille des hipposidéridés, *Palaeophyllophora quercyi* ; (3) une civette de moyenne taille, *Stenoplesictis minor* (famille des viverridés) ; (4) *Palaeotodus itardiensis*, todier de la famille des todidés ; (5) *Caenomeryx cf. procommunis*, petit herbivore artiodactyle de la taille d'un lapin, de la famille éteinte des cainothériidés. Au deuxième plan : (6) un grand oiseau carnivore, la mère et son petit, *Ameghinornis minor* (famille des phororhacidés) ; (7) *Stenogale gracilis*, petit carnivore du groupe des putois (famille des mustélidés) ; (8) *Idiornis itardiensis*, échassier voisin des cariamidés. Plus en arrière : (9) une famille de grands rhinocéros primitifs, *Ronzotherium sp.* (famille des rhinocerotidés) ; (10) *Nimravus intermedium*, grand carnivore félin à dents de sabre. Perché sur l'arbre, (11) un rolrier, *Geranopterus alatus* (famille des coraciidés).



2. UN CRÂNE ET UNE MANDIBULE d'un mammifère artiodactyle ruminant (le groupe qui renferme les vaches, les chameaux et les girafes) d'une quinzaine de kilogrammes, *Diplobune minor*, retrouvés dans la position même où ils étaient lorsque l'animal est mort, il y a 30 millions d'années. Le site d'Itardiès, d'où ils proviennent, a aussi livré des pattes du même animal en connexion anatomique. Leur analyse révèle que cette espèce était adaptée au déplacement sur des terrains boueux instables : l'environnement était sans doute marécageux. La qualité de préservation des fossiles du Quercy apporte ainsi de précieuses informations sur les parentés entre les espèces, mais aussi sur les évolutions de l'environnement pendant 35 millions d'années.

L'origine des chauves-souris

Dans l'ordre des chauves-souris, les Chiroptères, les fossiles du Quercy révèlent la diversification précoce, dès l'Éocène moyen (il y a 44 millions d'années), c'est-à-dire peu après son apparition, du genre *Stehlinia*, forme primitive de la superfamille des vespertilionoïdés, très répandue et très diversifiée dans la nature actuelle (et comprenant, entre autres, les chauves-souris que l'on rencontre aujourd'hui en France, tels les murins, les sérotines et les pipistrelles).

Le genre *Stehlinia* a été décrit en 1919 d'après un unique spécimen, un petit crâne (non daté) des anciennes collections des phosphorites du Quercy. Grâce aux fouilles modernes, des spécimens de ce genre ont été retrouvés, parfois assez complets, et datés avec précision. Un inventaire, établi en 1987, montre la coexistence de plusieurs lignées de *Stehlinia*, dans différentes catégories de taille, de l'Éocène supérieur à l'Oligocène supérieur (il y a entre 38 et 23 millions d'années).

Des *Stehlinia* plus anciens, de l'Éocène moyen (il y a 50 à 38 millions d'années), avaient été reconnus dans quelques faunes karstiques ou dans des dépôts lacustres d'Europe occidentale (en France, en Suisse et en Allemagne). En général, il s'agissait de dents isolées ou de rangées dentaires partielles, dont l'attribution à une espèce plutôt qu'à une autre était impossible. Seule la cavité de Cuzal, découverte récemment dans le Quercy, a livré des restes plus complets, qui ont permis à l'un d'entre nous (B. Sigé) et à Gerhard Storch de démontrer la coexistence de trois lignées dès l'Éocène moyen en Europe, qui perdurent entre 44 et 41 millions d'années.

Cette diversité, quasi contemporaine de la plus ancienne trace de *Stehlinia*, atteste la diversification adaptative précoce de ce genre en différentes catégories de taille. Cette diversification a-t-elle suivi de peu l'émergence en Europe du type morphologique *Stehlinia* à partir d'une forme plus primitive, ou est-elle consécutive à l'arrivée en Europe d'une forme plus ancienne apparue sur un autre continent ? Quelques éléments autorisent cette dernière hypothèse, notamment la présence en Tunisie de chiroptères anciens, de l'Éocène inférieur, ce qui montre la précocité de l'ensemble vespertilionoïde dans cette aire géographique.