

Data mining :

la loi ne doit pas enterrer la recherche française

Les Echos

Le 25 Avril 2016 paru à 15:48 et mis à jour à 16:45

Continuer à empêcher le TDM va mettre les chercheurs français dans une position de faiblesse vis à vis de leurs collègues étrangers

Au nom de la protection du droit d'auteur, la loi interdit aux chercheurs de réutiliser des articles scientifiques publiés afin d'en exploiter les données. Une entrave à la recherche qu'il faut supprimer.

En créant des freins juridiques multiples à la fouille massive des données et textes scientifiques (en anglais "TDM"), les députés et les sénateurs pourraient malgré eux handicaper gravement la recherche publique française pour les 10 ans à venir.

Alors que nous commençons à peine à entrevoir les perspectives ouvertes par l'accumulation et la fouille massives de données, tous les acteurs économiques et industriels ont mobilisé leurs ressources et leur intelligence pour explorer les avancées potentielles qu'elles permettent. Paradoxalement et tristement, la Recherche publique française est empêchée de rejoindre ce mouvement à cause de multiples obstacles juridiques.

La fouille de textes et de données (text and data mining - TDM) consiste à explorer, via un algorithme de fouille, des corpus immenses - composés notamment d'articles scientifiques ou de données expérimentales - afin d'en déduire des nouvelles connaissances. Imaginez qu'un chercheur en épidémiologie puisse analyser automatiquement les 2 millions d'articles scientifiques traitant du cancer, afin de déterminer s'il existe une corrélation inconnue à ce jour entre une pratique et l'apparition de la maladie : voilà ce que permet le TDM. Si cela vaut pour l'ensemble des disciplines, cette technique prend tout son sens lorsqu'il s'agit d'objets complexes, qui nécessitent un traitement pluridisciplinaire : les épidémies donc, mais aussi les objets climatiques, les phénomènes politiques, etc.

Si cela n'est pas permis en France, c'est que des dispositions du droit d'auteur et du droit sur les bases de données interdisent - sauf convention - la réutilisation des articles scientifiques... même par ceux qui les ont produits ! Cela est déconcertant car les chercheurs ont déjà payé l'accès sur ces articles. Rappelons en effet que les bibliothèques universitaires et les organismes de recherche dépensent plus de cent millions d'euros par an pour s'abonner aux revues scientifiques et permettre aux chercheurs d'y accéder. C'est précisément sur ces contenus que le TDM demeure impossible en pratique.

Continuer à empêcher le TDM va mettre les chercheurs français dans une position de faiblesse vis à vis de leurs collègues étrangers exerçant aux États-Unis, Canada, Japon, Grande-Bretagne... qui pourront utiliser le TDM grâce aux législations plus ouvertes et

permissives de leur pays.

Pour éviter de porter un coup qui pourrait être fatal à la recherche française, les pratiques de TDM doivent être enfin clairement autorisées et encouragées par les pouvoirs publics. Ce n'est qu'à cette condition qu'il sera possible d'éviter le départ de chercheurs français à l'étranger. Une véritable exception au droit d'auteur doit donc être inscrite dans le projet de loi pour une République numérique : la Commission européenne a d'ailleurs annoncé qu'elle allait défendre cette mesure dans le cadre du marché unique numérique. L'Allemagne s'apprête à réviser sa loi fédérale sur ces thèmes, sans attendre la révision préalable de la Directive européenne sur le droit d'auteur.

Il faut toutefois impérativement accompagner une telle mesure d'une politique audacieuse de recherche et de développement scientifiques capables de dépasser les limites du data mining tel qu'il est actuellement pratiqué – et qui consiste à porter à ses extrémités la calculabilité au détriment de l'interprétation théorique, comme l'a mis en évidence Chris Anderson dans *The end of theory*. L'Europe doit relancer une politique de recherche ambitieuse – celle qui fut l'origine du web - au lieu de se contenter de singer la Silicon Valley aujourd'hui, et peut-être, demain, la Chine.

Pour le Gouvernement et les parlementaires, faire le choix du TDM ne doit pas être vécu comme une entrave au droit d'auteur, mais bien comme un acte de soutien à la recherche et à la souveraineté scientifique de la France.

L'amendement qui a été déposé au Sénat, qui veut régler la question du TDM via les contrats entre éditeurs et chercheurs, est insatisfaisant. En effet, au vu des rapports de forces inégaux, les contrats de ce type sont défavorables aux chercheurs et compliquent considérablement leurs pratiques. À cet égard, le texte adopté par l'Assemblée Nationale reflète la volonté de la communauté scientifique. Il est issu des travaux et concertations engagées depuis plusieurs mois avec l'ensemble des acteurs de la recherche : le législateur ne se borne pas à renvoyer au contrat mais inscrit le cadre d'une véritable exception législative au droit d'auteur, au nom de l'intérêt général. Cette rédaction est à même de donner aux chercheurs français les moyens effectifs d'utiliser le TDM pour préserver le haut niveau de compétitivité de la recherche française. Si la voie contractuelle devait être tout de même retenue, il faudrait modifier cet amendement en spécifiant par décret les limitations techniques à prohiber. Ce n'est qu'ainsi qu'il sera possible de donner à la recherche française les moyens de continuer sa mission, qui permet notamment de nourrir l'innovation et donc la croissance.

Freiner le TDM, c'est empêcher la France de se saisir des nouvelles opportunités scientifiques et économiques créées par l'hyper-accélération des technologies de la donnée.

Collectif d'entrepreneurs et de dirigeants du numérique

- **Mounir MAHJOUBI, Président du Conseil national du numérique,**

- **Yann BONNET, Secrétaire général du Conseil national du numérique,**

Jean-Loup SALZMANN, Président de la CPU (Conférence des Présidents d'Université),

- **Alain FUCHS, Président du CNRS,**

- **Yves LEVY, Président directeur général de l'INSERM,**

- François HOULLIER, Président de l'INRA,
- Vincent BERGER, Directeur de la Recherche Fondamentale au CEA
- Antoine PETIT, Président Directeur Général d'INRIA et membre du Conseil national du numérique,
- Yves POILANE, Directeur de Télécom ParisTech et membre du Conseil national du numérique,
- Bruno CAHUDRET, Président du Conseil Scientifique du CNRS,
- Thierry DAMERVAL, Directeur général de l'INSERM,
- Christophe PERALES, Président de l'ADBU (Association des directeurs et personnels de direction des bibliothèques universitaires et de la documentation),
- François HOULLIER, Président de l'INRA ,
- Jean-Luc BEYLAT, Président de l'Association Française des pôles de compétitivité et President du pole Systematic Paris Region,
- Amal TALEB, Vice-Présidente du Conseil national du numérique,
- Godefroy BEAUVALLET, Vice-Président du Conseil national du numérique,
- Jean-Paul MOATTI, Président directeur général de l'IRD (Institut de Recherche pour le Développement),
- Benoit THIEULIN, Directeur de l'École de la Communication de Sciences Po, ex-Président du Conseil national du numérique (janvier 2013- janvier 2016) et actuellement membre du Conseil national du numérique,
- Gilles BABINET, Digital champion de la France auprès de la Commission européenne et ex-Président du Conseil national du numérique (avril 2011- décembre 2015),
- Jean-Marc BOURNIGAL, Président d'IRSTEA (Institut national de recherche en sciences et technologies pour l'environnement et l'agriculture),
- Emmanuelle JANNES-OBBER, Président de l'EPRIST (association des responsables de l'information scientifique et technique des organismes de recherche français publics ou d'utilité publique),
- Hélène JACQUOT-GUIMBAL, Directrice générale de l'IFSTTAR,
- Serge PIPERNO, Directeur scientifique de l'IFSTTAR,
- Michel EDDI, PDG du CIRAD,
- Bernard STIEGLER, Philosophe et Président d'Ars Industrialis (association internationale pour une politique industrielle des technologies de l'esprit),

- **Patrick BRAOUEZEC, Président de Plaine Commune,**
- **Renaud FABRE, Directeur de l'information scientifique et technique du CNRS,**
- **Benjamin JEAN, Président d'OpenLaw,**
- **Daniel KAPLAN, Co-fondateur et Délégué général de la Fondation pour l'Internet Nouvelle Génération (FING) et membre du Conseil national du numérique,**
- **Véronique TORNER, Co-fondatrice et Co-présidente d'Alter way et membre du Conseil national du numérique,**
- **Carole ZISA-GARAT, CEO de TELEGRAFIK et membre du Conseil national du numérique,**
- **Jean-Marc PATOUILLAUD, Managing Partner de Partech et membre du Conseil national du numérique,**
- **Emmanuelle ROUX, Co-fondatrice du FacLab de l'Université de Cergy-Pontoise et directrice-associée de Sc21 et membre du Conseil national du numérique,**
- **Marc TESSIER, Administrateur de VidéoFutur, Président du Forum des Images et membre du Conseil national du numérique,**
- **Rand HINDI, Fondateur de Snips et membre du Conseil national du numérique**