

Communiqué de presse

200 ans d'archives de presse ouvertes à tous grâce à un partenariat technologique entre *Le Temps* et l'EPFL

Lausanne, vendredi 18 mars 2016

Le laboratoire d'humanités digitales de l'Ecole polytechnique fédérale de Lausanne, *Le Temps* et la Bibliothèque nationale suisse s'associent pour lancer le 18 mars 2016, à l'occasion du 18^e anniversaire du *Temps*, la plus importante base de données historiques de Suisse. Grâce aux équipes du professeur Frédéric Kaplan, les archives du *Journal de Genève*, de la *Gazette de Lausanne* et du *Nouveau Quotidien*, remontant à 1798, sont accessibles à tous et consultables sur un site exploitant les technologies d'indexation les plus avancées pour les documents historiques.

«Ces archives constituent un système d'information à part entière qui nous permet de recréer dans le détail non seulement l'histoire suisse des deux cents dernières années, mais aussi une grande partie de l'histoire mondiale», explique Frédéric Kaplan.

Toute nouvelle recherche d'archive ne se limitera plus à énumérer les articles dans lesquels un mot apparaît, comme le font les moteurs de recherche classiques, mais indique l'évolution du nombre d'occurrences de ce mot au fil des ans. Cet aperçu chronologique permet à l'internaute de visualiser en un coup d'œil la fréquence d'un terme à travers plus de deux cents ans d'histoire et de se focaliser sur la période qui l'intéresse. Ce moteur de recherche ouvre de nouvelles perspectives tant aux professionnels qu'à un vaste public.

«Nous sommes ravis de pouvoir offrir un outil d'une telle qualité à nos internautes. La collaboration avec le laboratoire d'humanités digitales, leader dans son domaine, nous a ouvert des possibilités nouvelles dont nous n'aurions pas osé rêver», insiste Gaël Hurlimann, rédacteur en chef du digital au *Temps*.

50 millions de noms de lieux et de personnalités reconnues

Parallèlement à la recherche lexicale, l'équipe du laboratoire d'humanités digitales a conçu un outil d'extraction permettant d'identifier tous les lieux et les personnes dans les 4 millions d'articles de l'archive, ainsi qu'un certain nombre de caractéristiques les concernant. Une base de données de 50 millions de mentions d'entités est mise en ligne aujourd'hui et interrogeable à travers un outil de recherche sémantique. Il devient par exemple possible de demander au système toutes les fonctions occupées par une personne au fil des ans, ou de trouver tous les articles impliquant des intervenants exerçant des métiers ou fonctions spécifiques. Ce sont autant de requêtes impossibles à formuler lorsqu'on se limite à la recherche classique de mots. Cette immense base de données est elle-même interconnectable aux autres bases du Web sémantique. Elle s'enrichira par ces nouvelles connexions et viendra nourrir les autres bases, donnant à l'histoire suisse une place sans précédent dans des bases de données mondiales.

10 ans de travail

Le lancement de cet outil révolutionnaire signe l'aboutissement d'une décennie de travail. En 2006 débute la numérisation de deux ans d'histoire via les archives du *Journal de Genève*, de la *Gazette de Lausanne* et du *Nouveau Quotidien*, donnant vie aux archives historiques du *Temps*.

Les progrès scientifiques faits notamment en linguistique computationnelle amèneront l'EPFL et *Le Temps* à s'associer en 2013 pour développer à partir des archives historiques du *Temps* cette base de données historiques. Dès 2014, la Bibliothèque nationale suisse s'associe à son tour au développement de ce moteur de recherche, soutient financièrement le projet et travaille en coordination étroite avec le laboratoire d'humanités digitales de l'EPFL sur les nouveaux outils développés. Deux ans seront ensuite nécessaires pour extraire des 4 millions d'articles de presse numérisés les index qui sous-tendent les nouveaux outils proposés. Ces données sont dès ce jour accessibles. Celles du *Nouveau Quotidien* seront disponibles à fin mars.

Un corpus d'articles utile pour la recherche

Les étudiants et chercheurs du laboratoire d'humanités digitales exploitent les archives historiques du *Temps* dans leurs travaux scientifiques et planchent déjà sur d'autres applications publiques, au premier rang desquelles plusieurs projets se penchent en particulier sur une approche «big data» de ces archives. Ils cherchent par exemple à analyser l'évolution de la langue française à travers les années, la manière dont le traitement de certains sujets perdure ou se modifie dans l'histoire et inversement, ou les caractéristiques qui permettent d'automatiquement dater un texte selon son style et son contenu.

«Grâce à cette archive, les étudiants ont l'opportunité de développer des approches computationnelles ouvrant la voie à de nouvelles méthodes historiques. Elles sont également susceptibles de transformer la linguistique et l'étude de médias», précise Frédéric Kaplan.

Un partenariat large

Les nombreux défis de ce projet entamés en 2006 ont pu être surmontés grâce à plusieurs partenaires et sponsors.

La numérisation de plus d'un million de pages issues de journaux aux formats variés a été rendue possible grâce au soutien de : la Bibliothèque nationale suisse, la Bibliothèque de Genève, la Bibliothèque cantonale et universitaire de Lausanne, Sandoz – Fondation de Famille, la banque privée Mirabaud & Cie SA et PubliGroupe.

L'indexation sémantique est quant à elle née de l'immense travail du laboratoire d'humanités digitales de l'EPFL piloté par le comité scientifique responsable de la méthodologie.

Membres du comité scientifique

- Alain Clavien, Professeur, Histoire contemporaine, Université de Fribourg
- Marie-Christine Doffey, Directrice, Bibliothèque nationale suisse
- Gaël Hurlimann, Rédacteur en chef Digital, *Le Temps*
- Joëlle Kuntz, Journaliste, *Le Temps*
- Enrico Natale, Directeur, Infoclio.ch
- François Vallotton, Professeur, Centre des sciences historiques et de la culture, Université de Lausanne

Découvrir le nouveau site : letempsarchives.ch

Contacts

- Gaël Hurlimann, rédacteur en chef en charge du numérique, *Le Temps* et *L'Hebdo*

Mobile: + 41 79 570 14 76

- Frédéric Kaplan, Professeur, Laboratoire d'humanités digitales, EPFL

Mobile : +41 78 705 78 36