



AVIS DE PRESENTATION DE THESE EN SOUTENANCE POUR L'OBTENTION DU DIPLOME NATIONAL DE DOCTEUR

Monsieur Axel JEAN - CAURANT

Présentera ses travaux intitulés :

**« Analyse de documents et du comportement des utilisateurs pour améliorer l'accès à
l'information »**

Spécialité : **Informatique et applications**

Le 8 octobre 2018 à 9h00

Lieu :

**Université de La Rochelle
Pôle Communication, Multimédia et Réseaux
Amphithéâtre
44 Av. Albert Einstein
17000 LA ROCHELLE**

Composition du jury :

M. BURIE Jean-Christophe	Professeur, Université de la Rochelle
M. COURBOULAY Vincent	Maître de conférences, HDR, Université de la Rochelle
M. DOUCET Antoine	Professeur, Université de la Rochelle
Mme EGLIN Véronique	Professeure, INSA de Lyon
Mme IDMHAND Fatiha	Professeure, Université de Poitiers
M. ILLOUZ Charles	Professeur, Université de la Rochelle
M. PLAS Pascal	Professeur, Université de Limoges
M. XANTHOS Aris	Maître d'enseignement et de recherche, Université de Lausanne

Résumé :

L'augmentation constante du nombre de documents disponibles et des moyens d'accès transforme les pratiques de recherche d'information. Depuis quelques années, de plus en plus de plateformes de recherche d'information à destination des chercheurs ou du grand public font leur apparition sur la toile. Ce flot d'information est bien évidemment une opportunité pour les utilisateurs mais ils sont maintenant confrontés à de nouveaux problèmes. Auparavant, la principale problématique des chercheurs était de savoir si une information existait. Aujourd'hui, il est plutôt question de savoir comment accéder à une information pertinente. Pour résoudre ce problème, deux leviers d'action seront étudiés dans cette thèse.

Nous pensons qu'il est avant tout important d'identifier l'usage qui est fait des principaux moyens d'accès à l'information. Être capable d'interpréter le comportement des utilisateurs est une étape nécessaire pour d'abord identifier ce que ces derniers comprennent des systèmes de recherche, et ensuite ce qui doit être approfondi. En effet, la plupart de ces systèmes agissent comme des boîtes noires qui masquent les différents processus sous-jacents. Si ces mécanismes n'ont pas besoin d'être entièrement maîtrisés par les utilisateurs, ils ont cependant un impact majeur qui doit être pris en compte dans l'exploitation des résultats. Pourquoi le moteur de recherche me renvoie-t-il ces résultats ? Pourquoi ce document est-il plus pertinent qu'un autre ? Ces questions apparemment banales sont pourtant essentielles à une recherche d'information critique. Nous pensons que les utilisateurs ont le droit et le devoir de s'interroger sur la pertinence des outils informatiques mis à leur disposition. Pour les aider dans cette tâche, nous avons développé une plateforme de recherche d'information en ligne à double usage. Elle peut tout d'abord être utilisée pour l'observation et la compréhension du comportement des utilisateurs. De plus, elle peut aussi être utilisée comme support pédagogique, pour mettre en évidence les différents biais de recherche auxquels les utilisateurs sont confrontés.

Dans le même temps, ces outils doivent être améliorés. Nous prenons dans cette thèse l'exemple de la qualité des documents qui a un impact certain sur leur accessibilité. La quantité de documents disponibles ne cessant d'augmenter, les opérateurs humains sont de moins en moins capables de les corriger manuellement et de s'assurer de leur qualité. Il est donc nécessaire de mettre en place de nouvelles stratégies pour améliorer le fonctionnement des systèmes de recherche. Nous proposons dans cette thèse une méthode pour automatiquement identifier et corriger certaines erreurs générées par les processus automatiques d'extraction d'information (en particulier l'OCR).