
I. 1.1 SÉRENDIPITÉ DE L'INFORMATION

La sérendipité ou "serendipity" est un mot qui a vu le jour en 1754 sous la plume du philosophe anglais Sir Horatio Walpole, dans son roman "The three princes of Serendib" (les trois princes de Ceylan). Ces princes jouissaient de cet insigne privilège de la providence qui consistait à « tomber » systématiquement par hasard sur la bonne solution, quelque fâcheuse que fut la situation dans laquelle ils se trouvaient. Par ce terme, Walpole désignait la faculté de certaines personnes à trouver la bonne information par hasard sans réellement l'avoir recherchée. Le grand dictionnaire du Québec définit la sérendipité comme la « Rencontre, au cours d'une observation empirique, de données ou de résultats théoriquement inattendus, aberrants et capitaux »¹.

Si l'histoire des princes de Serendib est pour une large part oubliée, le mot, lui, est resté. Avec les sciences cognitives et l'avènement du web il a même connu une seconde vie. Originellement, précise le Grand Dictionnaire, ce terme « a été introduit dans l'usage scientifique par le physiologiste W.B. Cannon (1945) et surtout, en ce sens précis, par le sociologue R.K. Merton (1951, trad. fr. Mendras, 1953), à partir d'un terme épistolaire de H. Walpole (28 janv. 1754) ». Désormais, dans le contexte de la recherche d'information on entend par sérendipité, une heureuse rencontre du hasard et de la nécessité, qui, par la grâce de l'intuition comme de la capacité d'inférence de l'utilisateur, lui permet de trouver et par là même de compiler, les bonnes informations sur un sujet donné, participant ainsi directement au processus d'élaboration de la connaissance. De ce fait, la sérendipité concourt à la « maîtrise de l'entropie informationnelle »².

Pour [Ertzscheid & al. 2003] la sérendipité en recherche d'information correspond à « la propagation d'un style cognitif stable (mis en place au début de la session de navigation) dans un environnement différent mais contenant de l'information pertinente pour l'usager dans le contexte initial de sa navigation et au vu de la tâche qu'il s'était assigné ». Constatant qu'une recherche ainsi initiée donne des résultats satisfaisant son besoin d'élaboration de connaissance, l'utilisateur va poursuivre et affiner sa stratégie de recherche d'information sur la base du même « style cognitif ». La notion de sérendipité ainsi définie peut être alors vue notamment comme le processus de

¹ Source : www.granddictionnaire.com

² Ertzscheid, O., Gallezot, G. – « Chercher faux et trouver juste » - X^e Colloque bilatéral franco-roumain, CIFSIC Université de Bucarest, 28 juin – 3 juillet 2003, Bucarest : (2003) - http://archivesic.ccsd.cnrs.fr/sic_00000689/fr/

reformulation progressive de la requête par l'utilisateur, dans les systèmes de relevance feedback interactifs.

Le schéma proposé par [Ertzscheid, 2003] correspond sur le plan des processus mis en œuvre, aux trois schémas « initiaux » des processus de recherche d'information.

Etat initial	Processus	Modèles
[Je sais] [ce que je cherche]	Querying / Browsing Sérendipité nulle	Computational
[Je ne sais pas] [ce que je cherche]	Searching Sérendipité structurelle	Utilisateur
[Je sais] [que je ne sais pas ce que je cherche]	Learning Sérendipité associative	Environnementaliste

FIGURE 1 - ERTZSCHEID & AL :
SCHEMAS « INITIAUX » DES PROCESSUS DE RECHERCHE D'INFORMATION

Le premier cas correspond à la recherche d'information classique. Lorsque l'utilisateur sait ce qu'il cherche, il met en place une stratégie fondée sur des mots-clés qu'il sait a priori pertinents et lance la recherche à partir de ces mots-clés. La logique mise en œuvre est alors celle du « matching ». Dans le second cas, l'utilisateur raisonne par déduction et par abduction car il ne sait pas ce qu'il cherche. C'est le cas de recherche ouverte ou exploratoire. Le système dans ce cas précis a pour fonction d'amener l'utilisateur à trouver ce qu'il ne connaît pas³. Le dernier, cas, lorsque l'utilisateur sait qu'il ne sait pas ce qu'il cherche, le schéma d'utilisation du moteur de recherche est alors celui qui met le plus en jeu le phénomène de sérendipité. L'utilisateur est alors en situation de mettre en œuvre une stratégie de d'intuition-déduction-approfondissement, quelle que soit la complexité du système utilisé. On se trouve alors dans une « situation « d'apprentissage périphérique » comme l'ont défini [Lave & Wenger, 91]⁴. Il s'agit d'une stratégie typiquement liée à la navigation hypertextuelle sur internet ou sur une base documentaire. L'utilisateur individuel se trouve alors en situation de « découverte » plus que de recherche, de l'information [Boutin & al. 2007]⁵. L'utilisateur, même professionnel ne possédant pas la maîtrise des algorithmes de recherche d'information, est contraint d'accepter une part d'aléatoire, étrange composite de pertinence, hasard et

³ Belkin N., Helping People Find What They Don't Know, in Communications of the ACM, August 2000, Vol. 43, No. 8.

⁴ Lave G., Wenger E., Situated Learning: Legitimate Peripheral Participation. New-York, Cambridge University Press, 1991.

⁵ Boutin, E Ertzscheid, O., Gallezot, G : « Perspectives documentaires sur les moteurs de recherche : entre sérendipité et logiques marchandes » - article en cours de soumission. (2007) - http://archivesic.ccsd.cnrs.fr/sic_00172169/fr/

inexactitudes, dans la restitution des résultats d'un système de recherche d'information. Cette part d'aléatoire ou sérendipité systématique « apparaît ainsi comme une composante majeure de cette médiation qu'est le processus de recherche d'information » [Boutin & al. 2007].

I. 1.2 NOTION DE PERTINENCE CONTEXTUELLE.

Eric Boutin⁶ [Boutin 2005] évoque la notion de biais cognitif pour évaluer la pertinence contextuelle d'un document par rapport à une requête. Plusieurs idées viennent modifier la notion classique de pertinence qui consiste en l'identification de mot-clés de la requête à l'intérieur des documents renvoyés par un moteur de recherche. La requête, dans cette optique, est d'abord considérée comme un désir d'en savoir plus sur un sujet : elle devient dès lors vecteur d'élaboration de connaissance. Cette analyse se situe dans une logique d'édification de la connaissance à partir des informations progressivement compulsées par l'utilisateur. Ce qui ressort de ces travaux est notamment que le lecteur se concentre prioritairement sur les premières pages qu'il consulte, accordant de ce fait un poids supérieur aux informations contenues dans ces pages. Pour l'usager, ce sont les premiers documents consultés qui ancrent la représentation dominante de sa recherche de connaissance. D'autre part, la notion de pertinence vue par l'utilisateur ne s'accommode pas de la redondance. La pertinence d'un tel document est considérée comme nulle. Il s'agit alors de redondance négative. La notion de pertinence est donc corrélée à la notion de nouveauté d'une information qui vient enrichir un champ de connaissance.

Carbonell et Goldstein⁷ proposent une vision renouvelée de la pertinence qui combine pertinence et anti-redondance et introduisent la notion de pertinence marginale maximale. Cette vision combine des mesures de pertinence textuelle par rapport à une requête et le degré de nouveauté d'un document par rapport à l'ensemble des résultats renvoyés par le moteur. Ainsi un document aura un taux élevé de pertinence marginale maximale, en fonction de sa capacité à répondre à la requête de l'utilisateur d'une part et d'autre part s'il est suffisamment différent des documents précédents.

⁶ BOUTIN, E. (2006). Biais cognitifs et recherche d'information sur internet. Quelles perspectives pour les indicateurs de pertinence des moteurs de recherche. In Actes du colloque VSST, Janvier 2006. Lille, France.

⁷ Jaime Carbonell, Jade Goldstein, "The use of MMR, diversity based reranking for reordering documents and producing summaries" ACM Conference on Research and Development in Information Retrieval, Actes du colloque p 335-336, 1998.

[Boutin 2006] estime qu'il est plus difficile pour un individu de modifier une idée acquise ou préconçue que d'élaborer une nouvelle connaissance pour laquelle il ne possède pas encore de modèle ou schéma mental. Aronson⁸ cité par Boutin estime que la « dissonance cognitive » est proportionnelle à l'effort fourni pour sera d'autant plus forte, que l'effort aura été élevé pour poser la première pierre de l'élaboration d'un champ de connaissance.

[Boutin 2006] introduit la notion de pertinence conditionnelle, car les documents, au regard de l'utilisateur ont une pertinence liée à l'ordre dans lequel ils ont été consultés dans la mesure où « 80% des connaissances issues de recherche d'information se forgent à partir des 10 premiers résultats renvoyés par le moteur de recherche ». D'autre part, la notion de pertinence vue par l'utilisateur ne s'accommode pas de la redondance. La pertinence d'un tel document est considérée comme nulle. La notion de pertinence est donc corrélée à la notion de nouveauté d'une information qui vient enrichir un champ de connaissance. Mais la redondance n'est pas toujours négative. Elle peut également servir à confirmer et donc valider les connaissances acquises lors de la consultation des premiers documents, et sert ainsi à établir de nouveaux schémas mentaux et ancrer une connaissance nouvellement acquise dans le champ des possibles. La redondance devient alors positive. [Benyu & a 2005]⁹ proposent deux métriques importantes au regard de la notion d'élaboration de connaissance. Il s'agit des métriques de diversité et de richesse. Ces deux métriques ont pour base le sens de la requête de l'internaute qui est vue comme une question posée au système. La diversité correspond à la quantité d'informations retournées par un moteur de recherche et la richesse mesure l'exhaustivité des réponses apportées à la question de l'internaute. « La pertinence des pages est alors construite à l'issue d'un processus calculatoire original, directement transposé du Pagerank de Google mais appliqué à une matrice des similarités textuelles entre documents plutôt qu'à une matrice des relations hypertextuelles entre pages web ».[Boutin 2006]. Cette vision de la notion de pertinence basée sur l'élaboration de connaissances en contexte, semble être une alternative originale aux métriques classiques de précision et de rappel utilisées depuis 1992 pour évaluer les systèmes de recherche d'information notamment lors de la conférence TREC.

⁸ ARONSON E., (1973) "The Rationalizing Animal," Psychology Today, May 1973, pp. 46-51.

⁹ BENYU Z.; HUA L.; YI L.; LEI J.; WENSI X.; WEIGUO F.; ZHENG C.; WEI-YING M., (2005), Improving Web Search Results Using Affinity Graph, conference WWW 2005

I. 2 L'APPROCHE ABDUCTIVE

L'abduction fait appel aux qualités de déduction et de perception des utilisateurs. Elle suscite la capacité de raisonnement des usagers des systèmes de recherche d'information. Il s'agit d'un processus cognitif capital dans la démarche de construction de connaissance, qui consiste à adopter temporairement des « hypothèses susceptibles d'être vérifiées ultérieurement » [Cattelin 2004]¹⁰. L'abduction n'est pas un acte en soi mais une pratique d'ordre exploratoire, à mettre au rang des habitus, qui sous-tendent à la fois des réflexions et savoirs sous-jacents intervenant dans le processus d'élaboration de connaissance. Elle témoigne de la capacité d'adaptation de l'être humain face à une situation inconnue. De ce point de vue là, l'abduction, appliquée à la recherche d'information est le processus cognitif qui figure à la source d'une démarche d'essai-erreur-apprentissage et concourt de ce fait à l'élargissement ou au contraire à la précision progressive d'un champ de recherche, permettant ainsi à un utilisateur d'accumuler des connaissances hétéroclites sur un sujet ou, au contraire, d'atteindre précisément une cible.

Ainsi une analyse des processus inférentiels pourrait-elle permettre d'évaluer la part d'interaction entre un utilisateur et un système de recherche d'information, en situation de démarche d'essai-erreur apprentissage. Une piste d'estimation logique du processus abductif est suggérée par [Ertzscheid & al. 2003]¹¹ dans un article consacré à la sérendipité de l'information et que nous évoquons par ailleurs plus en détail à la section 6.1.1, consacrée à ce sujet.

Une estimation du processus abductif peut être approchée grâce aux quatre équations de Figueiredo et Campos [Figueiredo, 01] cités par [Ertzscheid & al. 2003]. Ces équations sont fondées sur une analyse de la nature métaphorique de la requête. La métaphore est alors utilisée comme instrument de stimulation de la perspicacité de l'utilisateur.

Les auteurs distinguent les éléments suivants : Le problème (P), le contexte du problème (KP), la métaphore (M), le contexte de la métaphore (KM), la solution (S), le contexte de la solution (KS) et l'élargissement du champ de connaissance dans le processus de formulation du problème (KN). Quatre types de métaphores sont envisagés : la métaphore simple, la métaphore inattendue, l'absence de métaphore, la métaphore de l'ignorance.

¹⁰ Sylvie Cattelin : « L'abduction : une pratique de la découverte scientifique et littéraire » in Critique de la raison numérique -

¹¹ Ertzscheid, O., Gallezot, G. - « Chercher faux et trouver juste » - X^e Colloque bilatéral franco-roumain, CIFSIC Université de Bucarest, 28 juin - 3 juillet 2003, Bucarest : (2003) - http://archivesic.ccsd.cnrs.fr/sic_00000689/fr/

Dans la première équation la métaphore simple amène l'utilisateur à trouver une solution.

$$\begin{array}{l} P1 \subset (KP1) \\ M \subset (KM) \end{array} \Rightarrow S1 \subset (KP1, KM, KN)$$

FIGURE 2 - EQUATION N°1 - EXEMPLE D'ARCHIMEDE - METAPHORE SIMPLE

Dans la deuxième équation la métaphore inattendue propose un nouvel éclairage qui amène l'utilisateur à lancer une seconde recherche avant d'aboutir à une solution par abduction simple.

$$\begin{array}{l} P1 \subset (KP1) \\ M \subset (KM) \end{array} \Rightarrow \begin{array}{l} P2 \subset (KP2) \\ S2 \subset (KP2, KM, KN) \end{array}$$

FIGURE 3 - EQUATION N°2 - METAPHORE INATTENDUE - INFERENCE INDUCTIVE

Dans la troisième équation, l'absence de métaphore fait appel au processus d'inférence abductive, qui permet de remonter le fil de l'information par une démarche d'essai-erreur amenant, de requête en requête à trouver une solution.

$$P1 \subset (KP1) \Rightarrow \begin{array}{l} P2 \subset (KP2) \\ S2 \subset (KP2, KN) \end{array}$$

FIGURE 4 - EQUATION N°3 - INFERENCE ABDUCTIVE SIMPLE

Dans la quatrième équation, la métaphore de l'ignorance introduit l'erreur dès la formulation de la requête. Cette requête erronée amène un nouveau problème, puis, par un processus progressif plus long que le précédent, une nouvelle solution.

$$P1 \subset (KP1, EP1) \Rightarrow \begin{array}{l} P2 \subset (KP2) \\ S2 \subset (KP2, KN) \end{array}$$

FIGURE 5 - EQUATION N°4 - INFERENCE ABDUCTIVE COMPLEXE