

DESCRIPTIF DES DONNEES

Tableau 1: présentation synthétique des 3 expériences

Expériences	VI manipulées	Facteurs additionnels	N	Type d'établissement
Etude 1 : Comment les adolescents recherchent-ils des informations sur Google en situation scolaire ? Impact de la complexité et de l'expertise sur les comportements rationnels et la décision d'arrêter la recherche	- Connaissances antérieures (V.I. inter-sujet, domaine de connaissances : Histoire) - Complexité des tâches de RI (V.I. intra-sujet, 3 niveaux : TC, TR, TI)	- SAE - Connaissances antérieures	38	Collège
Etude 2 : Régulation métacognitive de la RI sur internet : le rôle du FOF (feeling-of-findability) et des connaissances antérieures sur les calculs de taux de gains dans le cadre de la décision d'arrêt	- Connaissances antérieures (V.I. inter-sujet, domaine de connaissances : Biologie Appliquée) - Complexité des tâches de RI (V.I. intra-sujet, 3 niveaux : TC, TR, TI)	- SAE - Connaissances antérieures - Mesures métacognitives (FoF)	29	Lycée Professionnel
Etude 3 : Effet de la complexité des tâches et des connaissances antérieures sur l'heuristique combinée : exploration de l'indépendance des deux règles cognitives	- Connaissances antérieures (V.I. inter-sujet, domaines de connaissances : SVT vs ECO) - Complexité des tâches de RI (V.I. intra-sujet, 3 niveaux : TC, TR, TI)	- SAE - Connaissances antérieures - Mesures métacognitives (FoF)	42	Lycée Général

Mesures relatives aux comportements de recherche

Parmi les indicateurs retenus pour investiguer les interactions avec le moteur de recherche et les comportements d'analyse des utilités espérées sur les pages SERP (discrimination des liens et choix des documents), ont été mesurés :

- **Nombre de documents ouverts.** Nous avons compté toutes les pages Web ou documents (pdf, vidéo, image, etc.) avec une URL différente qui ont été ouverts. Ce nombre de documents sera spécifiquement utilisé dans les analyses comportementales de l'étude 1 et dans toutes les études pour le test de l'heuristique. Pour rappel, nous défendons l'idée que chaque document ouvert relève de la phase d'exploitation (accumulation de l'information pertinente).
- **Temps d'exploration.** Nous avons mesuré le temps passé sur les pages de résultats du moteur de recherche (i.e. le temps passé sur les pages Google entre chaque document, incluant le temps de formulation/reformulation, considérant le temps d'exploration comme un temps dédié à la construction du chemin d'accès à l'information ou temps entre les patchs dans l'IFT – voir Chapitre 2 pour plus de détails).

- **Temps d'exploitation.** Nous avons mesuré le temps passé sur les pages et documents Web consultés à effectuer le traitement de l'information contenue, depuis l'ouverture en cliquant sur le lien jusqu'au départ de la page Web. Pour rappel, le temps d'exploitation est considéré dans cette thèse (et dans le cadre de l'IFT – voir Chapitre 2 pour plus de détails) comme le temps de l'accumulation de l'information pertinente.

Mesures relative au compromis exploration-exploitation

La RI est une activité de régulation exploration-exploitation, chaque phase représentant une séquence précise d'actions. Dans cette perspective, les transitions entre ces deux phases d'exploitation et d'exploration manifestent des calculs d'utilité spécifiques (Fu & Pirolli, 2007 ; Pirolli, 2005 ; Pirolli & Card). Pour analyser ces transitions et mesurer les seuils d'utilité conditionnant les actions de navigations, 2 types de transitions ont été mesurées :

- **Le nombre de transitions SW2a.** Comme le montre le modèle 1d (schéma 7), le nombre de transitions SW2a représentent le nombre de retours sur une page SERP effectués depuis un document ouvert par un utilisateur. Ces transitions indiquent d'une part l'inflexion des rendements sur le document visité (gains d'informations ou utilité espérée devenus plus faibles que l'utilité moyenne incitant à la sortie), et d'autre part la confiance de l'utilisateur dans l'utilité du prochain lien sur la SERP à permettre la poursuite de l'accumulation de l'information pertinente.
- **Le nombre de transitions SW2b.** Comme le montre le modèle 1d (schéma 7), le nombre de transitions SW2b représentent le nombre de reformulations effectuées par un utilisateur. Ces transitions indiquent d'une part que les utilités espérées des liens ou des documents proposés sur la page SERP sont trop faibles (gains d'informations ou utilité espérée plus faibles que l'utilité moyenne incitant), et d'autre part le manque de confiance de l'utilisateur dans l'utilité de la requête précédemment formulée, dont la page SERP est dans notre thèse un substitut (voir Chapitre 2 pour plus de détails). Ainsi, l'utilisateur fait le calcul que les prochains liens sur la SERP ne permettront pas de trouver l'information manquante (i.e. la poursuite de l'accumulation de l'information pertinente) et est incité à reformuler. Le nombre de SW2b se calcule donc par :

$$SW2b = Reformulations = (Nombre\ total\ de\ requêtes - 1)$$

Mesures relative à la performance des utilisateurs

- **Le temps de complétion ou temps total allouée à la tâche de RI.** Nous avons mesuré le temps mis par les participants pour réaliser la tâche de RI, i.e pour trouver l'information manquante et construire la réponse.
- **Le taux d'information pertinente accumulé** (taux de gain total). Cette mesure est obtenue en rapportant le gain réel total d'information pertinente au temps passé à l'exploitation (temps à lire ou étudier le document), en utilisant la formule de l'IFT proposée par des études antérieures (Pirolli & Card, 1999, Pirolli, 2007, 2007 ; Nakayama & Wan, 2021) tel que $\pi = g/T_w$, avec g le gain net et T_w , le temps d'exploitation (temps à l'intérieur du patch, voir chapitre 2 pour plus de détails). Soit : $\pi = (score)/(temps\ d'exploitation)$. Plus ce taux est élevé, plus le participant a su accumuler d'information pertinente par unité de temps consacrée à la tâche (meilleure absorption de l'information, plus grande efficacité de la collecte). Plus ce taux est élevé, plus la vitesse d'accumulation (collecte de l'information pertinente) est élevée.

Note 1 : toutes les mesures de temps ont été effectuées en milliseconde, puis ont subi une log-transformation avant utilisation.