The impact of expertise on query formulation strategies during complex learning task solving: A study with students in medicine and computer science

Cheyenne Dosso1, Lynda Tamine2, Pierre-Vincent Paubel1 & Aline Chevalier1
1University of Toulouse, CILLE-CNRS, 31058 Toulouse, France – cheyenne.dosso@univ-tlse2.fr
2University of Toulouse, IRIT-CNRS, 118 Route de Narbonne 31062 Toulouse, France

Introduction

During information search (IS) activity with a search engine:
- Users have to plan their activity, evaluate information, deep process it and navigate between Search Engines.
- Results Pages (SERPs) and online documents
- Prior domain knowledge (PDK) allows users to:
  - Formulate more and longer queries
  - Produce more keywords related to the domain knowledge
  - Several results were obtained in the past about the interaction between IS/PDK when users solve information search problem, few recent studies focus on these variables in the case of Search as Learning.
- Search engine becomes an information tool allowing users to become learners during IS activity.

To prior domain knowledge sustains search behavior for formulating queries to solve information search problem, it is relevant to study the impact of PDK when users solve complex learning tasks versus information problem such as:
- Information problem
  - Simple task (fact-finding goal)
  - Inferential (production of new keywords more relevant than those provide in the statement)
- Complex learning tasks
  - Exploratory learning (searching about a topic)
  - Making-decision (comparison of various elements provided in the statement)
- PDK solving (creation and elaboration of a new solution/production)

According to theoretical background, this study aimed to:
- To better understanding how users formulate queries depending on their level of prior domain knowledge according to the search and learning task complexity
- To better understanding users’ difficulties for reformulating queries while solving complex learning tasks with regard to their level of domain expertise

Methods

Participants:
- Ten students in Computer Sciences
- Ten students in Medicine

- Six males / 4 females
- Age – M: 23.8 SD: 3.2
- Master degree / PhD
- SAE information search; M: 33.5 SD: 5

- Five males / 5 females
- Age – M: 25.4 SD: 3.4
- Master degree / PhD
- SAE information search: M: 30.8 SD: 4.83

- Self-report 5-point Likert scale of PDK – Computer Sciences domain: M: 4.2 SD: 0.42
- Self-report 5-point Likert scale of PDK – Medical domain: M: 4.1 SD: 0.5

- Self-report 5-point Likert scale of PDK – Computer Sciences domain: M: 2.5 SD: 1.2
- Self-report 5-point Likert scale of PDK – Medical domain: M: 1.3 SD: 0.7

- Computer Sciences knowledge test (PDK assessment); M: 6.8 SD: 2.2
- Computer Sciences knowledge test (PDK assessment); M: 0.5 SD: 0.7

- Medical knowledge test (PDK assessment); M: 1.5 SD: 0.7
- Medical knowledge test (PDK assessment); M: 5 SD: 0.7

Data Analyses

For each dependent variable, we performed an ANOVA (repeated measures) on the three independent variables:

IV
Level of prior domain knowledge as between-subject factor: (high/low)
Type of task as within-subject factor: (simple/exploratory/learning/decision-making/problem solving/inferential)
Domain of task as within-subject factor: (medicine/computer sciences)

Results

DV1a – Total number of new keywords produced and DV1b – Query length: No significant (p > .05)
DV2 – Total number of keywords used from tasks:

PDK: No significant (p > .05)

DV3a – Total number of new keywords produced by users from specific vocabulary

PDK: No significant (p > .05)

DV3b – Total number of new keywords produced by users from specific vocabulary

PDK: Total number of new keywords produced by users from specific vocabulary

PDK*Domain

PDK*Type of task and PDK*Type of task*Domain: No significant (p > .05)

Conclusion

Main Results:
- When participants have lower prior domain knowledge, they used keywords from task statements more often than participants with higher prior domain knowledge.
- This effect was particularly true for decision-making and inferential tasks.
- Computer Sciences produced more new keywords from general vocabulary (i.e. words from common language) in computer sciences tasks than in medicine tasks.
- Medicine users tended to formulate more queries with more domain specific words related to a high vocabulary in medicine than Computer Sciences users when solving medicine tasks.

Explanation:
- Users with higher prior domain knowledge modified easier learning task goals to others terms, whereas users with lower prior domain knowledge needed to rely on words from statements.
- The generation of new keywords tended to be less specific in computer science and more specific in medicine.
- Computer science vocabulary (e.g. software, programming...) are words fallen in the everyday language whereas medicine words are more specific to this domain

Limits and further work:
- To study only query formulation strategies is not enough to understand relationships between search behavior and learning
- To bring deeper result interpretations, we will analyze the relevance of the outcomes (answers) provided as well as variables from questionnaires completed before and after each tasks (e.g. expected and perceived difficulty, self-perception of answer quality)
- Further studies should investigate prior domain knowledge on search abilities during Search as Learning.
- The aim is to link search and learning variables to determine difficulties experienced by users without PDK and how users with do to perform better than users without PDK during complex learning tasks

References


The French National Research Agency (ANR), CoST-Modelling Complex Search Tasks (ANR-18-CE33-0016), supported this research.