A "Searchable" Space with Routes, for Querying Scientific Information

Renaud Fabre

April 4, 2019

A "Searchable" Space with Routes for Querying Scientific Information

Renaud Fabre

Laboratoire Paragraphe (EA349), Université Paris 8, Saint-Denis, France
renaud.fabre01@gmail.com

Abstract. Users searching for scientific information are confronted by a "hidden face" of searchable space: their own selection of items, which could help map their navigation on recorded search "routes", is not open to consultation, and remains either concealed or even unavailable. At best, users benefit from recommendations but, as data on their own choices are not shared, networked information for global navigation remains nebulous.
This position paper tackles the following research question: How could users searching for scientific information benefit from each other's search sessions? Our answer comprises two steps. First, using examples from our own data, we look at the characteristics of user behavior, approached here mainly via structures of collaborative personalized search. The second step proposes a mapping of recorded search sessions: for similar queries, search sessions are modeled by sets of typical user/item pairs in networks we call "eco-systems" of queries. These eco-systems connect search sessions and are open to navigation for users. Final output is threefold: 1) visualization of interconnection between search sessions, 2) localization of the search session that best suits a user's needs, 3) navigation on alternative "routes" between search sessions selecting alternative paths to answers. The conclusion raises questions about typical uses of this new kind of a documentary object, exploiting functionalities originally adapted from a bipartite graph.

Keywords: knowledge management · search session . query modeling · information retrieval. bi-partite graphs

1 Introduction

Scientific and technical information (STI) has been recently qualified as a "complex system" that has to improve its accuracy and its digital organization: these opinions, which appear in the research agenda (The National Academies of Science and Medicine 2017) of the National Academies of Sciences of the United States, underpin this position paper which raises questions about "searchability" of STI.

"Searchable" means "capable of being computationally searched", within an independent community of search, open to serendipity (Conrad and Moeller, 2017), and navigation in an unknown space, even if we could now consider that: "Web search is governed by a unified hidden space, and each involved element such as query and URL has its inborn position" ([3]). In the STI community, searching is a highly dynamic practice as the "aims and intentions of the user's initial search evolve as new information is encountered" ([17]).

The main issue tackled by this position paper is to approach search sessions for any query as "communities" of information retrieval, which are together confronted by the "hidden face" of their common searchable space. Users of search sessions never receive a record of other users' generated contents. There is no mapping of the various results of search sessions, showing clusters of user selections of items: these data, even anonymized, are not open to consultation, and remain either concealed or even unavailable. The lack of these kinds of data significantly affects STI searches, where a single keyword could give access to a very wide range of structured discussions and interpretations, each delivered with their own specific "version" and "vocabulary" or field of experiment. Because of this limitation on alternative documentary selections, it remains difficult to identify roots of controversies: accurate interpretation of views, downloads, citations, is uncertain, and searching can sometimes appear as a lonely walk in a forest of hazy homonyms.

However, current research barely addresses the issue, as only "few approaches take advantage of searches performed previously by users" ([10]). Fortunately, recent approaches on interactive information retrieval and user behavior highlight that STI searches cover a large variety of distinct methods and needs ([12]). Based on the latter positive standpoint, in short, this article tackles the following research question: "How could STI search users benefit from each other's search sessions?"

Our answer is twofold. First, we look at the variety of search behaviors and the need for a clearer approach to networks of search sessions: for similar queries, search sessions could differ significantly in their selections of items according to users' variable needs, which are still hardly modeled. Second, we propose a mapping of search sessions with the help of a bipartite graph that we developed for navigation inside what we call the "eco-system" of answers to a query. The final point of this article is to review issues of efficiency relating to this new documentary object, modeling alternative searches, and helping to select, among recorded options, the answer that best fits a query.

2 Part 1. Search session and users' behaviors

STI search strategies are revealed by behaviors of the searches performed by users. These behaviors could be identified from structured "relations between the topics and the use of documents" ([4]), and users could choose between alternative strategies of search and use ([6]).

We will look at the characterization of user behavior in two areas: first is the community behavior of searches, which appears to differ among disciplines;

second, we review elements of individual behaviors which appear to vary with wide-ranging reference chasing purposes. Both seem to interact.

2.1 Community behaviors: variations among disciplines

To start with examples, differences in STI use behaviors among disciplines could be considered via the large-scale survey of research data management carried out at the CNRS (Centre National de la Recherche Scientifique) in 2014 ([16]). Results, which are part of a nationwide survey on scientific information uses,[1] include the opinions of 432 directors of French public research laboratories, answering 91 questions, out of 1,250 directors of laboratories who were asked to answer. Overall, with a rather high answer rate (above 30%), data reveal significant differences in STI uses among various fields of scientific practice. For instance, concerning adoption of a clearly identifiable common STI practice like Research Data Management (RDM), "we can distinguish three groups: (1) laboratories from nuclear and particle physics and from social sciences and humanities appear globally more advanced regarding RDM than other disciplines; (2) laboratories from the three domains of ecology and the environment, informatics, and earth sciences and astronomy have dedicated resources and make their data available; (3) laboratories in the field of physics appear aware of the challenge." Regarding Open Access practices, results also reveal significant differences in the management of STI data: these results all have direct impacts on the methods and purposes of search activities and on searched items.

Another type of community behavior appears in a further recent national survey (COPIST): also at the CNRS, we carried out a survey of STI management at the national scale for all research organizations. Results include answers from 106 research and higher education institutions and events, overall showing a strong desire to share STI digital practices and fairly large differences between current uses and management of groups of institutions. Detailed results on bibliometric uses and strategies are available:[2] they reveal significant differences in users' practices at the various institutions, together with their desire to share uses and data.[3]

2.2 Individual Behaviors: variations in reference chasing

It is established that emerging results create the risk of a "cold start" biased answer to a query: new content is difficult to retrieve as it has, by definition, not yet been produced elsewhere. Conversely, a versatile answer could be seen as fruitful when trying to find out items which could be identified together as "typical" and "original". It can also be the case that the attempted search

[1] http://www.cnrs.fr/dist/z-outils/documents/Enqu%C3%AAte%20DU%20-%20DIST%20mars%202015.pdf

[2] http://www.cnrs.fr/dist/z-outils/documents/copist-premiers-resultats.pdf, p. 55 and further

[3] *Ibid*, pp. 11 and 19

result is bound to unpredictable instrumental bases. [7] These search intentions could explain how "multiple search strategies" have been experienced in an STI context ([5]), and that a multiplicity of search tactics exist: up to 29 separate search methods have been described (Bates, 1979).

Search sessions can thus have numerous goals, be grounded in variable motivations, and differ along with variable discovery methods ([7]). Searching thus develops its own rationality: it could be said that "search is not research". Overall, the search context remains nebulous, with a permanent threat of information overload ([8]), and with questionable performances of recommender systems ([19]). User behavior modeling is not evenly covered by research: the field of observational ([9]) data is a significant example where a wide range of behaviors "appear to balance breadth and specificity," while authors clearly observe a wide range of differences between uses of reviewed disciplines. Modeling of search sessions meets with many open challenges, which are all based on "interactive IR" and the need to model it.

3 Part 2: Networking search sessions: a "Query Eco-System"

3.1 Interactive IR: a network for search sessions

Interactive IR proposes networking of search sessions in various type of frameworks like a "*collaborative query management system*" for "search and browse interaction" ([13]); it could be tested on data analyses of parsing on publishers' knowledge bases (PKB) or researchers' documentary logs. A better knowledge of user behaviors results in "collaborative personalized search"[20] which "satisfies the various information needs of different users" of the same query. In the practice of personalization techniques, a modeling system like PECIRS [15] gives examples of "a new user-centric mechanism". Personalization aims to "adapt search results to enhance the relevance to users according to their past search behaviors."

In this context, the "query ecosystem" presented here takes an innovative approach to personalization techniques, aiming at an "interactive information retrieval process" ([5]), interlinking changes in search results with changes in knowledge ([12]). Here personalization does not refer to users' past behavior but to visualization of all mapped search sessions and to autonomous selection of the search session that best suits a user's needs.

In a "query ecosystem", interaction of a set of search sessions corresponds to specific IR needs in given STI contexts: an example is when research infrastructures using the same kind of analysis source (X Rays, Neutrons, etc.) for a large variety of experiments, each with their own practices in the same science, produce comparable search sessions as "versions" of the same scientific object. In these contexts, the main goal of a query eco-system is to "re-rank results based on the user's intention" by formalizing an analyzed interaction between similar queries and retrieved results, and thus "lower the cognitive burden" of search by mapping search sessions on a documentary topic of common interest ([21]).

To build up representation of search sessions, we use graphs which have long been used for a large variety of tasks related to knowledge representation. Bipartite graph characteristics ([14]) are exploited in this article, with a typical *crown-graph* figure, familiar to neural network approaches, but used here in a novel manner, which we developed following ideas suggested by the information analysis capacities of graphs ([11]) tied to their typical geometry. To the best of our knowledge, there is no modeling of search sessions analysis comparable to the "query eco-system."

3.2 Definition 1: query routes and search sessions

Let us consider that our goal is to record "query routes" for a given keyword, and that these routes, which of course could differ, are designed to be compared ([2]).

Let us then write:

$$Q_n = f(N_n, K_n)$$

where Q_n is a number of queries produced by users of a browser. Let us also consider that for each query Q, we can record a number of users N and a number of items (URL, documents, articles) K recording the search for a keyword at the same time among the same corpus of items: we will call this group a community of users of the same query "route" in a search session .

For positioning any distinct route, we could express the limits of its system of "typical" search sessions ([10]) for a given query. Let us pose two limits of substitution between "users" and "items" for any query: one limit is where a few users ask for many items, and the other limit is when many users consult a very small number of items. We then could write these limits of possible substitution between users and items, N and K:

$$Q_1 = f(N_{\max, K_{\min}} \quad \text{or} \quad Q_1 = f(N_{\min, K_{\max}}$$

Or, in a general form:

$$\begin{gathered} [\max,\ \min \\ Q_n = \mathrm{fN}, K \\ [\max,\ \min \end{gathered}$$

3.3 Definition 2: variation of queried search sessions

Let us now note α the coefficient of increase of N and β the coefficient of increase of K when Q varies by one unit, that is to say, when a new additional query on the same keywords is recorded (for instance: check users interested in items of a domain before and after publication of a famous article). The above-mentioned coefficients of variation will allow measurement of the "stability" or instability of the search session's content, according to variations in coefficients of the increase or decrease of quantities n and k between two queries of the same "route". We will then record whether or not users and/or items have varied in a correlative way, between queries Q_1 and Q_2.

We assume here that the data structure of any query could vary from one to the other, and that this variation could be expressed by an observable and recordable index of change in the relation between number of users and number of items.

We could then write:

$$Q_n = f\ [\overset{\alpha}{\text{nmax, min}}\ \overset{\beta}{\text{kmax, min}}$$

Let us note that value "min" or "max" of n and k provide data on the measured quantity of these variables to "produce" Q, but let us also note that, from one unit Q to the other (from one query to the other), the coefficient of increase or of decrease between value "min" or "max" of n and k provides additional data on the quantity of these variables in the production of Q. These additional data provide a value of the "stability" or "instability" of the process of querying an additional search session, with a dynamic index of behavior. For any unit n and k in production of Q, the variation of quantities measured by α and β is critical information about Q.

3.4 Definition 3: networks of queried search sessions

We know that with regard to the slope of a set of queries, with the added mix of user and items that it measures, the slope tends to be stable when $\alpha + \beta \leq 1$ or $(\alpha + \beta) \rightarrow 1$: in the latter case, transition between Q_{n-1} and Q_n is steady because, as measured by $\alpha + \beta$, variation of $(n + k)$ is still of the same order as the variation of Q. There is no significant increase or decrease in the number of units (users and items) recorded between these two or more successive queries and corresponding search session content.

Conversely, there could be also an alternative situation in which $(\alpha+\beta)\pm\infty$, i.e. that for each variation of one unit of Q, we have an unstable variation of characteristics of Q, indicated by an unstable variation of $(\alpha + \beta)$, which will increase or decrease abruptly: relations between users and items will change strongly. In the latter case, between units Q_{n-1} and Q_n, the variation of the quantities x and y would be unstable, which means that the query's user and item pair will change significantly. This situation could indicate that a threshold has been reached in the effectiveness of the query.

In this case, why continue to allocate (n, k) to Q if $(\alpha + \beta)$ tends toward minus infinity $-\infty$? This would mean that the combination (n, k) becomes about "performing", which is questionable when the time comes to decide on a new query Q_{n+1} with the same characteristics of n and k. With such characteristics for a query, it could be more interesting to change n or k than to reproduce the same quantities in the next query. More generally, for transition between Q_{n-1} and Q_n, the choice of stability in n and k could be "*justified*" if we observe that stability prevails with $(\alpha + \beta)$ tending to 1 and "*questionable*" when instability prevails with $\alpha + \beta$ tending to $\pm\ \infty$.

On a long series of queries, there will be "learning" phenomena which will give meaning to the expression of limits in variations of the purposed investigation on a long set of queries covering large fields. In that perspective, it is possible to fix arbitrary limits of variations to a given set of queries, with the "best" and the "worst", for instance:

$$\text{Best} = N^{\alpha}_{\max}, (\alpha + \beta) \to 1, K^{\beta}_{\min}$$

And

$$\text{Worst} = N^{\alpha}_{\min}, (\alpha + \beta) \pm \infty, K^{\beta}_{\max}$$

If we have selected these two opposite limits to queries belonging to an STI data search, we could write our limits with the following framework:

"Best"	"Worst"
$\mathrm{N}^{\alpha}_{\max}$	$\mathrm{N}^{\alpha}_{\min}$
$(\alpha + \beta) \to 1$	$(\alpha + \beta) \pm \infty$
β	β
$K_{\min}$	$K_{\max}$

test

3.5 Definition 4: a "compass" positioning search sessions on routes

On the basis of the preceding expression, the two triplets could give shape to a formal graph-based presentation of *paths existing between these two limits* and possessing the characteristic of positioning all the *non-contradictory solutions* existing between these two limits with three parameters. The following bipartite Hamiltonian graph with connected nodes offers that representation.

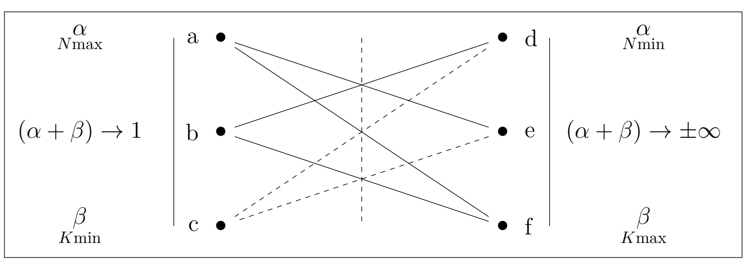

Fig. 1. "Compass" GRAPHYP: Positioning Typical Search Sessions

Figure 1 shows a complete representation of all the typical intermediary situation between our two limits, which gives us a tool for classification of observed queries Q in series of searches on a given keyword, according to user and item choices. Structured by GRAPHYP, this set of query search typical position has two main characteristics, which will be detailed below with the help of Table 1.

As observed in Table 1, GRAPHYP expresses the whole set of *non-contradictory positions that structured search routes could occupy during any search* on a given keyword in a database. This has consequences, defined by three characteristics, on GRAPHYP operating rules.

NODE (Query)	TYPICAL DIRECTION OF SEARCH (Users, Items)
a	$\mathbf{N}\alpha_{\ \mathbf{max}}$, $\mathbf{K}\beta_{\ \mathbf{max}}$, $(\alpha + \beta) \to \pm\,\infty$
b	$\mathbf{N}\alpha_{\ \mathbf{min}}$, $\mathbf{K}\beta_{\ \mathbf{max}}$, $(\alpha + \beta) \to \mathbf{1}$
c	$\mathbf{N}\alpha_{\ \mathbf{min}}$, $\mathbf{K}\beta_{\ \mathbf{min}}$, $(\alpha + \beta) \to \pm\,\infty$
d	$\mathbf{N}\alpha_{\ \mathbf{min}}$, $\mathbf{K}\beta_{\ \mathbf{min}}$, $(\alpha + \beta) \to \mathbf{1}$
e	$\mathbf{N}\alpha_{\ \mathbf{max}}$, $\mathbf{K}\beta_{\ \mathbf{min}}$, $(\alpha + \beta) \to \pm\,\infty$,
f	$\mathbf{N}\alpha_{\ \mathbf{max}}$, $\mathbf{K}\beta_{\ \mathbf{max}}$, $(\alpha + \beta) \to \mathbf{1}$

Table 1: Classification of Typical Search Sessions in GRAPHYP

Characteristic 1: GRAPHYP functions like a compass as it supplies all possible directions, and "locates" the recorded direction of a query Q. This set of possible "positions" during a search gives an overview of the proposed modeling of the searchable space for STI queries.

Characteristic 2: the second feature of GRAPHYP search modeling is that it can be used to record and compare search "behaviors" in querying. It allows users to *learn from their past recorded behavior* as well as from the recording of other users of the same base, if data is made accessible to all users. The anonymization solution is of course an important optional condition. [18]

In these two steps, the "compass" has the function of "orienting" and "locating" search attitudes as being recorded as individual dynamic behaviors in a search session.

Characteristic 3: GRAPHYP also allows the expression, by summation of individual attitudes, of the **global trend of a community** of users according to the characteristics of their uses in their own search sessions.

It could be possible to arbitrarily select one node, between a and e, to fix an **optimum point or optimum search session** there, and measure the distance of a set of observed routes toward the selected node. Mapping of STI search routes becomes possible, like in air or sea routes. Cooperative or conflicting routes could be identified by the GRAPHYP data structure.

3.6 Navigation Rule 1: a "progressive learning system" for search sessions

GRAPHYP creates opportunity to reach output (STI data search session best results) while identifying alternative ways to reach it for a given input (query). In this way, users' *learning profiles* express the way they access items and which items could be thus compared and located locally and globally in their search sessions. GRAPHYP could be used as a kind of a browser for mapping uses of search sessions for an STI community.

Like any compass, GRAPHYP secures navigational parameters at any scale (see infra), owing to its built-in characteristics of information processing and transfer, its functions for positioning previously recorded routes, and its capacity to record and recommend navigation paths. *GRAPHYP thus belongs to a new generation of learning systems combining intuitive choices and automatic path definition at any operational scale.*

3.7 Navigation Rule 2: locating search session routes

Let us take all nodes (a, b, c, d, e, f) as possible starting points for definitions of "goals" of routes of search sessions at time T. Any node could be designed as a "goal" to be reached on a route built from the starting point of a given other node. We could then record "search session profiles" corresponding to circulation from one node to the other on a recorded route between search sessions.

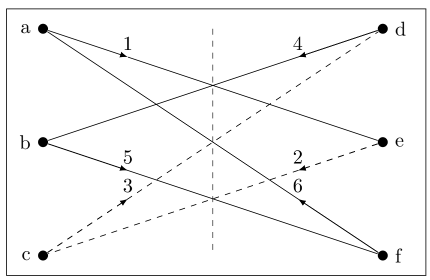

Fig. 2. Exploring Neighbor Search Sessions with GRAPHYP "compass"

From the design of Figure 2 above, we can observe that a continuous circulation from node to node shows a common edge in all the pairs corresponding to any node: node a has a common edge with node e, b has a common edge with f, etc. From this standpoint, we could remark that it is consistently possible to have circular "travel" inside GRAPHYP, from edge to edge, while coming back to the starting node by the way of the complementary edge of that node. In this way, search sessions could be interconnected and explored.

3.8 Navigation Rule 3: navigating between search sessions

As shown on Figure 2, node a could allow an exploration of GRAPHYP's other nodes which will represent six steps on one route, as designed here. Another six-step route could be practiced from the same node a starting from its other edge. There are at least 12 possible steps which could be explored from any node in the Eulerian circuit, and GRAPHYP thus offers 72 possibilities for identifying "search profile" characteristics which could be modeled and recorded in an STI query.

Numerous characterized positions in exploration of possible search sessions could be identified in this way: it offers users and their community detailed records of explored and unexplored ways of discovering information.

We could remark that these explorations of "possible" routes could be managed overall with any conventional probabilistic approach, with which the possibilistic approach could combine for any purposed uses.

3.9 Navigation Rule 4: fixing a search session profile

We could also apply the design of Figure 2 to propose another significant feature of traceability for searchable STI space, namely the recording of search profiles, used to "read" characteristic steps of a navigation route in a database. This result could be produced when projecting on x, y axes as shown in Figure 3.

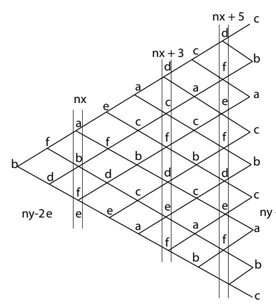

Fig. 3. Recording and Positioning of Nodes on Data Search Routes in GRAPHYP

Based on the design shown in Figure 3, we could, for any query applied to this grid of structured data (in our example: $nx + 3$ and $nx + 5$, $ny2$), identify the origin and thus the "kinship" of any search session located by GRAPHYP. We could then *evaluate all diverging or converging solutions surrounding any observed node located on the grid.* For any observed network, there exists an "**induced network**" which applies the property of "mutual reachability" of connected nodes in a network.

Mapping of search session routes using these ideas should help build strategies for discovering information.

3.10 Navigation Rule 5: mapping distances between search sessions

Exploring neighbor routes between search sessions is an important feature of the searchable space, as it provides information on its depth and treewidth and, as far as we know, there are hardly any existing developments in mapping of this aspect of searchable STI space.

Reducing uncertainty on the size and direction of future routes to discovery could benefit from representations of courses which have been followed and/or

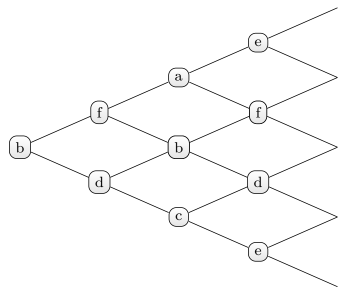

Fig. 4. Self-Similarity of the GRAPHYP Bipartite Graph

abandoned (Figure 3). Based on our example, exploration of neighbor arriving routes could benefit from tree couples of edges which shape this node:

$$a = (e, f) \quad b = (d, f) \quad c = (d, e) \quad d = (b, c)$$
$$c = (a, c) \quad f = (a, b)$$

Here we have the grid of nodes which are linked by a specified common edge, resulting from the position of each node on the grid. We could then propose the mapping of a grid of *correlated neighbor routes*. The grid observed on Figure 4, structured with the same nodes as the grid in Figure 3, offers the option of projection that could be studied in a further work.

3.11 Navigation Rule 6: fractal scalability of searchable space

A final property of the bipartite graph that gave birth to GRAPHYP is self-similarity, which is designed on the Figure 5 below.

The graph with nodes A, B, C, etc. is built on a larger dimension than the preceding one, and the addition of these "compasses" on a self-similar basis enables the building of information architectures of the same type for information processing at any scale, and the application to the GRAPHYP operating frame of basic addition, subtraction, multiplication and division operations.

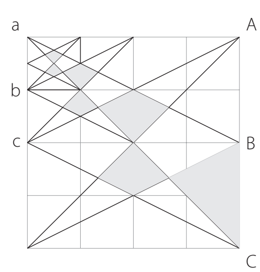

Fig. 5. Searching at Various Query Scales : Fractal Scalability of GRAPHYP

Scalability of the system could thus be established as a final feature of descriptive functionalities of the system. It could then be used as a tool connecting approaches at various steps or in different scientific domains.

4 Conclusions

The STI search community asks for accuracy and dynamic adaptation of delivered items, as highlighted by reviews of "observational" approaches to information retrieval ([9]), and this requires a rich semantic search ([1]).

We propose mapping of recordable categories of typical search sessions, clustered in user-item groups, which, together, shape what we called the "query ecosystem". Navigation in this system is organized for complete "searchability" of a query, modeling alternative paths to answers. It raises several issues: first, this new kind of a documentary object could help map "versioning" of search sessions, and thus add a new piece of semantic analysis for identification of alternative paths to answers, thus providing a "chronicle" of new incoming ideas through the corresponding modeling of searches. The status of sharing and privacy relating to this modeling system is a highly important issue.

Further work will seek to test accuracy and fast identification of versioning of search sessions; we will also check clustering at various scopes of fractal development of search session mapping, which would be useful for interactions between multipurpose searches (interdisciplinary, etc.) and the specific geometry of items networking inside this type of data structure.

References

1. Bast, H., Buchhold, B., Haussmann, E.: Semantic search on text and knowledge bases. Found. Trends Inf. Retr. **10**(2-3), 119–271 (Jun 2016). https://doi.org/10.1561/1500000032
2. Bendersky, M., Wang, X., Metzler, D., Najork, M.: Learning from user interactions in personal search via attribute parameterization. In: Proceedings of the Tenth ACM International Conference on Web Search and Data Mining. pp. 791–799. WSDM '17, ACM, New York, NY, USA (2017). https://doi.org/10.1145/3018661.3018712, http://doi.acm.org/10.1145/3018661.3018712
3. Bing, L., Zheng-Yu, Li, N.P., Lam, W., Wang, H.: Learning a unified embedding space of web search from large-scale query log. Knowledge-Based Systems, Elsevier (2018), available online
4. Cabanac, G., Chevalier, M., Chrisment, C., Julien, C.: Organization of digital ressources as an original facet for exploring the quiescent information capital of a community. Int J Digit Libr **11**, 239–261 (2010)
5. Carevic, Z., Lusky, M., van Hoek, W., Mayr, P.: Investigating exploratory search activities based on the stratagem level in digital libraries. CoRR **abs/1706.06410** (2017). https://doi.org/10.1007/s00799-017-0226-6, http://arxiv.org/abs/1706.06410
6. Fabre, R.: New Challenges for Knowledge: Digital Dynamics to Access and Sharing. Wiley (2016). https://doi.org/DOI:10.1002/9781119378112
7. Feyerabend, P.: Contre la méthode. Seuil, Paris (1975)

8. Gibney, E.: How to tame the flood of literature: Recommendation services claim to help researchers keep up with the most important papers without becoming overwhelmed. Nature **513**(7516) (2014)
9. Gregory, K., Groth, P., Cousijn, H., Scharnhorst, A., Wyatt, S.: “searching data: A review of observational data retrieval practices”. ArXiv (2017)
10. Gutierrez Soto, C.: Exploring the Reuse of Past Search Results in Information Retrieval. Ph.D. thesis, Université Paul Sabatier, Toulouse III (2016)
11. Haynes, P.S., Alboul, L., Penders, J.: Dynamic graph-based search inunknown environments. Journal of Discrete Algorithms (6 july 2011), online
12. Kacem, A., Mayr, P.: Users are not influenced by high impact and core journals while searching. GESIS (2018), bIR 2018 Workshop on Bibliometric-enhanced Information Retrieval
13. Khoussainova, N., Alazinska, M., Gatterbauer, W., Chul Kwon, Y., Suciu, D.: A case for a collaborative query management system. Cidr (2009), u. of Washington, Seattle
14. Kolda, S.G.: Measuring and modeling bipartite graphs with community structure. Journal of Complex Networks **vol 5**, 581–603 (2017)
15. Naderi, H., Rumpler, B.: Percirs: a system to combine personalized and collaborative information retrieval. Journal of Documentation **66**(4), 532–562 (2010)
16. Schöpfel, J., Ferrant, C., Andre, F., Fabre, R.: Research data management in the french national research center. Data Technologies and Applications (2018). https://doi.org/10.1108/DTA-01-2017-0005
17. Singh, V.: Ranking and query refinement for interactive data exploration. Procedia Computer Science **125**, 550–559 (2018). https://doi.org/doi.org/10.10 16
18. Smyth, B., Balfe, E.: Anonymous personalization in collaborative web search. Inf Retrieval **9**, 165–190 (2006). https://doi.org/10.1007/s10791-006-7148-z
19. Villegas, N., Sanchez, C., Daz-Cely, J., Tamura, G., Icesi, U., Colombia: Character-izing context-aware recommender systems: A systematic literature review. Knowledge-Based Systems **140**, 173–200 (2018)
20. Xue, G., Han, J., Yu, Y., Yang, Q.: User language model for collaborative personalized search. ACM Transactions on Information Systems **27**(2) (February 2009)
21. Yan, Z., Zheng, N., Ives, Z.G., Talukdar, P.P., Yu, C.: Active learning in keyword search-based data integration. The VLDB Journal **24**, 611–631 (October 2015)