A Survey of Definitions and Models of Exploratory Search

Emilie Palagi, Fabien Gandon, Raphaël Troncy, Alain Giboin

HAL Id: hal-01496557

https://hal.inria.fr/hal-01496557

Submitted on 27 Mar 2017

A Survey of Definitions and Models
of Exploratory Search

Emilie Palagi
Inria – Sophia Antipolis
2004 Route des Lucioles BP93
06902 Sophia Antipolis cedex
emilie.palagi@inria.fr

Fabien Gandon
Inria – Sophia Antipolis
2004 Route des Lucioles BP93
06902 Sophia Antipolis cedex
fabien.gandon@inria.fr

Alain Giboin
Inria – Sophia Antipolis
2004 Route des Lucioles BP93
06902 Sophia Antipolis cedex
alain.giboin@inria.fr

Raphaël Troncy
EURECOM
450 Route des chappes
06904 Sophia Antipolis cedex, France
raphael.troncy@eurecom.fr

ABSTRACT
Exploratory search has an unclear and open-ended definition. The complexity of the task and the difficulty of defining this activity are reflected in the limits of existing evaluation methods for exploratory search systems. In order to improve them, we intend to design an evaluation method based on a user-centered model of exploratory search. In this work, we identified and defined the characteristics of exploratory search and used them as an information seeking model evaluation grid. We tested this analytic grid on two information seeking models: Ellis' and Marchionini's models. The results show that Marchonini's model does not match our evaluation method's requirements whereas on the other hand Ellis' model could be adapted to better suit exploratory search.

CCS Concepts
H5.2 [Information interfaces and presentation]: User Interfaces --- Ergonomics, User-centered design, Evaluation/methodology.

Author Keywords
Exploratory Search; Evaluation; Exploratory Search's characteristics; Information Seeking models.

INTRODUCTION
Searching on the Web is not just about finding an answer to a specific question. In particular, exploratory search (ES) is performed whenever a user wants to discover a domain, increase his knowledge, learn about new topics, etc. Whatever the definition, characteristics and evaluation methods for ES are much less mature than their counterparts in classical search paradigms. In this paper we make a comparative study of the state of the art with regards to characteristics and models of exploratory search in order to lay the foundations for proposing criteria and methods for the design and evaluation of exploratory search engines.

The two research questions that we cover here are:

(1) According to the state of the art what characteristics are useful for defining exploratory search activity?

(2) What methodological elements can then be derived for the design and evaluation of an exploratory search process?

In section 2 we study the major contributions to the state of the art of the domain under consideration. In section 3, we identify and define the characteristics of ES. In section 4, we use these characteristics to analyze two information seeking models: Ellis' model and Marchionini's model.

RELATED WORK
Exploratory search definition
ES is a particular information seeking activity. It is a loosely defined concept as its definition is not stable and continues to evolve every time new systems are being developed. [10] was the first attempt to characterize what ES tasks are. Here, the author compares them to another well-known type of search activity: lookup search. Lookup is the most basic kind of search [2] and refers to focused searches where the user has a specific goal in mind and also an idea of the expected result. A typical example would be a user wanting to make a reservation to a restaurant and looking for the phone number on the Web. On the other hand, ES is described as open-ended, with an unclear information need (as in Belkin's anomalous state of knowledge [4]), an ill-structured problem of search with multiple targets. This search activity is evolving and can occur over time. For example, a user wants to know more about Senegal, she doesn't really know what kind of information she wants or what she will discover in this search session; she only knows she wants to learn more about that topic. Hence, the main goal in ES is learning. But, "learning in exploratory search is not only about memorization of salient facts, but rather the development of higher-level intellectual capabilities" [22].

Many papers use this dichotomy to define ES, maybe in an attempt to facilitate the understanding of this unclear concept. However, we believe that there is a continuum/spectrum between these two extremes, because we can find some lookup activities in an ES session [23]. Indeed, Marchionini depicts these two search activities as overlapping clouds, suggesting that "lookup tasks are

embedded in exploratory tasks and vice versa" [2]. In [2], the authors proposed a categorization between lookup and exploratory tasks according to two facets: the complexity (the number of paths involved in the search process) and the specificity of the search goal (precise or undefined), cf. **Table 1**.

Table 1. A categorization of Lookup and Exploratory search tasks [2]

	Low complexity	High complexity
Precise goal	Core lookup	Borderline lookup
Open ended goal	Borderline exploratory	Core exploratory

In that way, the line between lookup and exploratory searches activities is blurred.

Exploratory search is also related in the literature to other information seeking behaviors such as information foraging or berrypicking. The information foraging theory [17, 18] attempts to understand and explain how people seek information. The authors connected food foraging behaviors with information seeking behaviors, in the sense that behavioral patterns are similar. For example, based on the 'information scent' concept, information seekers detect and use cues (e.g. Web links or bibliographic citations [16]) to move from one information patch to another, looking for relevant information to their goal. We can find that kind of behavior also in ES.

Berrypicking model [3] is one of the first which highlights the exploratory nature of information seeking activity. In this work, Marcia Bates proposed a dynamic, nonlinear and evolving search process. She compares the information seeking process with the act of picking berries on bushes, where people must pick berries singly [23]. The point is, the searcher moves into the information space to find relevant information, one by one, dispersed into several documents. All along the way, every step gives new ideas to the searcher and can redefine the query or the search goal. It's a constant renewal of the information need. In an ES task, users also adopt a berrypicking strategy.

Information foraging and berrypicking involve complex interfaces and specific complex human-computer interactions. This is also the case for exploratory search systems. Information architect of these systems have to provide specific features to match the users' needs and behaviors.

Exploratory search models and evaluation

The complexity of the task and the difficulty to defining the ES activity are reflected in the limits of existing evaluation methods of exploratory search systems. The evaluation of exploratory search systems is recognized as a difficult and subtle activity because "it entails a qualitative and quantitative analysis both of the user behavior and of the search results" [5]. These complex systems combine several functionalities and behaviors forming an alchemy difficult to evaluate [15]. The assessment cannot be the same between classic search systems and exploratory search systems.

Some of exploratory search systems, like in [13], are evaluated with the standard Information Retrieval precision and recall metrics. However, these metrics are mainly focused on the result ranking

and they suppose a precisely identified search target and result set. They are not sufficient because the success of exploratory search systems does not only depend on the search algorithm [19, 23]. The overall design must be taken into account. Indeed, in [23] the interactions between the system and the user are considered as intentionally symbiotic. Furthermore in [14], Mark Nolan said that "in the design of search results and interfaces for browsing rich information resources we need to design a certain degree of elasticity into the product to give users more control over the results". This involves "highly interactive interfaces" [1] with whom users can actively perform their ES.

An exploratory search system evaluation method is always based on an ES process model. It may be explicit or implicit, detailed or succinct [15]. But the current evaluation methods do not exactly reflect the ES task. In [5], the authors extended Kuhlthau's Information Seeking Process model [9] and use it in their evaluation. However, the main problem in this model, and the reason it cannot reflect the entire process, is the exploratory component: ES is the main activity of the user, not a fraction of the entire search process [23].

In summary, evaluation methods of exploratory search are still incomplete. A better understanding of the user (e.g. her needs, her behaviors when she performs her ES task, etc.) should allow a more accurate assessment of exploratory search systems and, thereafter, a more effective improvement.

Our objective is to design a new evaluation method of exploratory search systems which will be based on a user-centered model of ES activity. To achieve this, despite the absence of an established ES's model, we have to analyze several information seeking process models by confronting them to the definition of ES in literature. Based on this, every model which is too far from the definition would be dropped.

Therefore, in the following sections we will first identify the characteristics of ES appearing most often in the description of this search activity. We will derive them for an evaluation process of classic information seeking models.

CHARACTERISTICS OF EXPLORATORY SEARCH

The description of the ES process is unstable. In our work which consists in designing a user-centered methodology of exploratory search systems evaluation, we are looking for an information seeking process model of ES. As exposed previously, the fact we base our methodology on an information seeking process model allows us to define users' ES behaviors and the kinds of tasks an exploratory search system intends to support. In this way, we want to analyze models to verify if they match the ES process. Even if the definition of ES is open ended, some characteristics can be frequently found in the literature and the different descriptions of this search activity.

In order to design an information seeking process models evaluation grid, we listed the characteristics that appear most often in literature. We also synthesized some of them (especially the serendipitous concept) so as to adapt them to ES or to facilitate the understanding.

Table 2. Characteristics of exploratory search

#	Characteristics	Definition	References (non exhaustive list)
1	An evolving search process	The user adopts an opportunistic behavior, and will change or specify the objective or goals of search or even the strategies used to achieve them through multiple queries reformulation or refinement. During the search, the user can accomplish forward or backward steps.	[11, 13, 15, 22–24]
2	Several one-off pinpoint searches	Throughout the search session, the user can do several one-off pinpoint searches, e.g. she's looking for a specified information to better understand a result or the reason why it was proposed. These pinpoint searches can be related to the exploratory search task or not. This is closely related to sensemaking activities.	[10, 22, 23]
3	An evolving information need	Throughout the search session, the user has an evolving information need. The elements or results discovered may change her information need and the way she first considered the framework of the search. This evolution of the information need may appear several times in one search session. It is closely related to characteristic n°1.	[11, 22–24]
4	Multiple targets/ goals of search	The user may not have one single precise goal, but rather one vague objective and several smaller goals which may change or evolve during the exploratory search task so as to achieve it.	[2, 11, 12, 15, 22, 23]
5	Multiple possible answers	As the user has one vague objective and several smaller goals to achieve it (see characteristic n°3.), the user might not have one precise answer but an aggregate of relevant information which will help her go further in her reflection and exploratory search process.	[2, 11, 12, 20, 23]
6	Not an expected exact answer		
7	A serendipitous attitude	It is the faculty to be surprised and to pay attention to it. The user carries out her search by adopting a serendipitous attitude; with such open mindedness, she can allow herself to be surprised by one unexpected element. She then exploits this discovery by changing the search strategy or search goal/objective, etc.	[10, 20, 22]
8	An open ended search activity which can occur over time	The user might never end her exploratory search. She can stop it for multiple reasons (she considers she has enough information to perform another task for example; she doesn't have time to carry on the search; etc.), and she will continue the search few hours/days/weeks/months/years later.	[10, 12, 15, 22–24]
9	An Anomalous State of Knowledge (ASK) and an ill-structured (vague, general or unsure) context of search or goals	At the beginning, the user has an ASK and a general context of search: she knows the motivation to start the search, but does not have a precise idea of what she is actually looking for (type of results, kind of information). She only has a lack of knowledge, a vague objective of search but no specific of definitive plan to attain it.	[11, 15, 22–24]
10	Multifaceted	During the exploratory search, the user selects one or multiple filters or *facets*, to explore the information space. She will try to find an approach to her problem, she may find an angle of attack or a framework which may include these facets of the explored subject.	[2, 11, 22–24]
11	Uncertainty is fluctuating	The user starts the search with an intense feeling of uncertainty. The level of uncertainty is intrinsically linked to the specification of the problem. The further the user goes in her search tasks (she will specify her objective and maybe define an approximate plan), the more she reduces her uncertainty. But if somewhere along the way she changes her objectives, the uncertainty will tend to increase again.	[11, 23, 24]

An open ended search activity which can occur over time

An evolving search process

ES TASK

Multifaceted

Several one-off pintpoint searches

Multiple possible answers

Uncertainty is fluctuating

Multiple targets/ goals of search

An evolving information need

Not an expected exact answer

USER

An Anomalous State of Knowledge (ASK) and an ill-structured (vague, general or unsure) context of search or goals
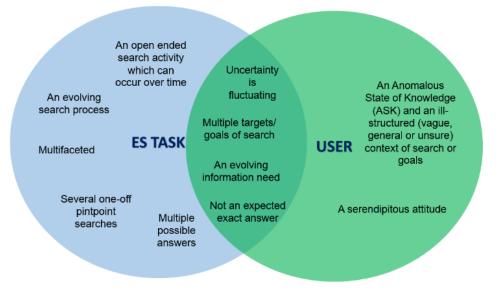
A serendipitous attitude

Figure 1. Venn diagram of characteristics of exploratory search

Defining characteristics

Most of the ES characteristics we found mentioned in the literature come from the same few references: [10, 23, 24].

Even though the same characteristics can be found elsewhere under different descriptions, they refer to the same idea. We summarized and listed these characteristics in **Table 2**.

The expression "serendipitous attitude" that describes the user's state of mind, cannot be found in the literature. Nevertheless, this idea or concept, matches other descriptions / characteristics such as *serendipitous discoveries* or *opportunistic behavior*. We were inspired by Sylvie Catellin's description of the serendipity concept [6].

In **Table 2**, the order of characteristics is not meaningful. All are equal insofar as the whole characteristics reflects the ES behavior. Furthermore, some of them are linked, e.g. characteristics 5 and 6 do not refer to the same thing but they share the same definition because they are linked. Moreover, some of them refer to others characteristics, e.g. the first characteristic, the *evolving process*, refers in its definition to characteristics 3, 4, 5, 6 and 7.

Characteristics' clustering

Based on the previous descriptions of the ES's characteristics, we clustered them into two categories (see **Figure 1**):

- The **User** category refers to characteristics which are related to the inner state of the user (feelings, attitude, expectations, etc.).
- The **ES task** category refers to characteristics which are related to the ES task and its process, *e.g.* the description of the task or strategies employed by the user.

Figure 1 divides the eleven characteristics into three groups: (1) the characteristics about the ES task only, (2) the ones about the inner state of the user and (3) those which refer to both the description of the search task and the user. It also depicts the imperative need to take into account the user in the exploratory search system development process: from the design to the assessment. We have to understand user's specific needs and the task she performs to adapt and improve exploratory search systems.

EVALUATION PROCEDURE AND PRELIMINARY RESULTS OF AN ONGOING WORK

The next stage is to analyze several information seeking process models and confront the author's description of the information seeking process model to our list of characteristics of ES. Although we can conceive that any model would not completely fit the ES definition, the idea here is to emphasize models that best satisfy the ES characteristics. If we can adapt one of them to match with the ES characteristics, it may be an imperfect or incomplete ES model, but good enough to help us in the design of an evaluation method of exploratory search systems.

The closest model to these characteristics will be the one we will select in our exploratory search systems evaluation method.

Methodology

We are continuously looking for ES's characteristics for each information-seeking process model using **Table 2** as an analytic grid. For a selected model, we check if the characteristic: (1) is explicitly mentioned in the description provided by the author(s); or (2) can be inferred from the description; or (3) is absent or cannot be inferred. Note that the specification of the inferences may show the possibility we will have to adapt the model, in order to have a model which covers all the ES's characteristics listed previously.

Results of the checking are reported in **Table 3**; in this table, case (1) is coded as _Yes_ (the characteristic is present), case (2) is coded as _Yes (Inferred)_, and case (3) is coded as _No_ (the characteristic is absent).

Preliminary test of the procedure and characteristics on two information seeking models

As we said before, we intend to apply the ES's characteristics grid on several information seeking models. We chose to restrict our interest because information seeking takes more into account the

user in the search process than information retrieval models which are more focused on the technology, the algorithms, etc.

In [10], Marchionini's model proposes among other things a set of activities related to ES and highlighted those associated to exploration (learn and investigate) or lookup. This model presents the interplay between these activities (and sub-activities). This seminal model is often used by authors in an exploratory search thematic.

Ellis' model (in its 1993 extended version) [7, 8]) is based on results of interviews of academic scientists on their information seeking activities. This model did not take a diagrammatic form and Ellis did not consider the different behaviors as a set of stages: he rather preferred *features* or *characteristics*. This model is well-known in the discipline and several authors based their own model on Ellis' one modulo an adaptation or modification.

To complete our analytic grid of models, we looked into authors' model description.

To the ES's characteristics we added the notion of "*sequential model*" because we need a sequential model of the ES process divided in several steps in order to devise an exploratory search systems evaluation method. Indeed, we want to identify the different steps of an ES task during an exploratory search system assessment. A sequential model would help us to verify if the evaluated system supports every step of the ES task.

Table 3. Two information seeking models analysis with our ES's characteristic grid

Characteristics	Ellis' model	Marchionini's model
1	Yes (inferred)	Yes
2	No	Yes
3	No	Yes
4	No	Yes (inferred)
5	Yes (inferred)	Yes (inferred)
6	Yes (inferred)	Yes (inferred)
7	No	Yes
8	No	Yes
9	Yes (inferred)	Yes
10	No	No
11	No	No
12 (*Sequential model*)	No	No

Table 3 shows the result of the analysis of Ellis and Marchionini's models. Neither of the models checked all the ES's characteristics of our analytic grid. As we can see, Marchionini's model is the one which fulfils the most criteria. Nevertheless, this model cannot be referred to as a sequential model, and it is moreover impossible to adapt it in this way.

Indeed, Marchionini's model (2006) "proposed a set of search activities associated with an exploratory search process" [22] such as *knowledge acquisition*, *comparison*, *analysis* or *evaluation*. It does not correspond to the multi-step process we are looking for, and we need lower-level activities to mobilize in our method in user tests.

Concerning Ellis' model, the grid shows that information is lacking to fulfill the whole grid. [23] note that "most situations involving information seeking can be characterized by the Ellis model". Based on **Table 3** and the inferences we made, we can imagine that the model could be adapted to ES. Indeed, unlike the previous model, this one is composed of eight steps ("features") such as *starting*, *browsing* or *verifying*. However, this model did not define the interactions and interrelationships between the eight features. The authors suggested that the order of the features can vary, as an evolving process.

CONCLUSION

The definition of ES is still unclear and ill-defined. The exploratory search systems assessment cannot be the same as for classic search engines. Evaluating these systems imply a better understanding of the ES behavior. Indeed, we need to know what kind of tasks these systems intend to support. In this approach we want to design a user-centered methodology based on an ES process model.

In order to find the model on which our method will be based, we first identified the characteristics of ES used in the literature and defined them. Then, we used these characteristics as an evaluation grid to analyze information seeking models, and we tested it on Ellis' and Marchionini's models. The results showed that Marchionini's model doesn't match the requirements that the method we want to design imposes. On the other hand, Ellis' model could be adapted to support ES.

The ongoing work detailed here should be further pursued by analyzing other models such as Kuhlthau's Information Seeking Process model, or Wilson's model. It would also be interesting to analyze with the grid we proposed other models inspired by Ellis', such as Waterworth and Chignell's information exploration model [21] for example.

Moreover, we want to connect each ES's characteristics to exploratory search engine features. In this regard, we will be able to use them as guidelines for the design of exploratory search systems, and as a heuristic grid for their assessment.

REFERENCES

[1] Ahn, J. and Brusilovsky, P. 2013. Adaptive visualization for exploratory information retrieval. *Information Processing & Management*. 49, 5 (Sep. 2013), 1139–1164.

[2] Athukorala, K., Głowacka, D., Jacucci, G., Oulasvirta, A. and Vreeken, J. 2015. Is exploratory search different? A comparison of information search behavior for exploratory and lookup tasks. *Journal of the Association for Information Science and Technology*. (Nov. 2015), n/a-n/a.

[3] Bates, M.J. 1989. The design of browsing and berrypicking techniques for the online search interface. *Online Review*. 13, 5 (May 1989), 407–424.

[4] Belkin, N.J. 1980. Anomalous states of knowledge as a basis for information-retrieval. *Canadian Journal of Information Science-Revue Canadienne Des Sciences De L'Information*. 5, MAY (1980), 133–143.

[5] Bozzon, A., Brambilla, M., Ceri, S. and Mazza, D. 2013. Exploratory search framework for Web data sources. *The VLDB Journal*. 22, 5 (Oct. 2013), 641–663.

[6] Catellin, S. 2014. *Sérendipité: du conte au concept*. Seuil.

[7] Ellis, D. 1989. A behavioural approach to information retrieval system design. *Journal of Documentation*. 45, 3 (Mar. 1989), 171–212.

[8] Ellis, D., Cox, D. and Hall, K. 1993. A comparison of the information seeking patterns of researchers in the physical and social sciences. *Journal of Documentation*. 49, 4 (Apr. 1993), 356–369.

[9] Kuhlthau, C.C. 1991. Inside the search process: Information seeking from the user's perspective. *JASIS*. 42, 5 (1991), 361–371.

[10] Marchionini, G. 2006. Exploratory search: from finding to understanding. *Communications of ACM*.

[11] Marie, N. 2014. *Linked data based exploratory search*. Nice.

[12] Marie, N., Gandon, F., Giboin, A. and Palagi, É. 2014. Exploratory search on topics through different perspectives with DBpedia. *Proceedings of the 10th International Conference on Semantic Systems* (2014), 45–52.

[13] Nguyen, T.-N., Dinh, D.-T. and Cao, T.-D. 2015. Empowering Exploratory Search on Linked Movie Open Data with Semantic Technologies. (2015), 1–8.

[14] Nolan, M. 2008. IA column: Exploring exploratory search. *Bulletin of the American Society for Information Science and Technology*. 34, 4 (Sep. 2008), 38–41.

[15] Palagi, E. 2015. Quelle méthode ergonomique élaborer pour évaluer les moteurs de recherche exploratoire? *CORIA 2015* (2015).

[16] Pirolli, P. 2007. *Information foraging theory: adaptive interaction with information*. Oxford University Press.

[17] Pirolli, P. and Card, S. 1999. Information foraging. *Psychological review*. 106, 4 (1999), 643.

[18] Pirolli, P. and Card, S. 1995. Information foraging in information access environments. *Proceedings of the SIGCHI conference on Human factors in computing systems* (1995), 51–58.

[19] Qu, Y. and Furnas, G.W. 2006. What do the attributes of exploratory search tell us about evaluation. (2006), 51–53.

[20] Sack, H. 2015. The Journey is the Reward - Towards New Paradigms in Web Search. *Business Information Systems Workshops*. Springer International Publishing. 15–26.

[21] Waterworth, J.A. and Chignell, M.H. 1991. A model of information exploration. *Hypermedia*. 3, 1 (1991), 35–58.

[22] White, R.W. 2016. *Interactions with search systems*.

[23] White, R.W. and Roth, R.A. 2009. Exploratory Search: Beyond the Query-Response Paradigm. *Synthesis Lectures on Information Concepts, Retrieval, and Services*. 1, 1 (Jan. 2009), 1–98.

[24] Wildemuth, B.M. and Freund, L. 2012. Assigning search tasks designed to elicit exploratory search behaviors. *Proceedings of the Symposium on Human-Computer Interaction and Information Retrieval* (2012), 4.