

Human Factors in Forum Exploration: A Sensemaking Perspective

Fan Yang-Turner, Lydia Lau

School of Computing
University of Leeds
Leeds, United Kingdom
{f.yang-turner, l.m.s.lau}@leeds.ac.uk

Abstract. Increasingly, technology-driven tools are being built with the aim to help people make sense of big data on the web. To effectively support sensemaking, understanding the human factors in data exploration will better inform the design of those tools. This paper proposes an approach of evaluation which starts with an articulation of characteristics in human sensemaking from theories and models. A data exploration tool that claims to support sensemaking could then be evaluated against these characteristics. An application of the proposed approach is illustrated in this paper by an experimental study on a forum exploration tool. This approach has provided benefits in terms of guiding the evaluation design and examining the impact of technical features of the tool. This preliminary work is an initial step for evaluating sensemaking tools with a general set of questions that can address essential human factors.

Keywords: data exploration, forum, semantic web, sensemaking, human factors, scenario, analysis, design, evaluation

1 Introduction

Every day, Web users are facing more and more information available in larger, ever more rapid flows. The skill of making sense of large scale information is no longer required just of specialists in intelligence analysis, but increasingly of everyone [10]. People are constantly engaging in sensemaking which is “a motivated, continuous effort to understand connections (which can be among people, places, and events) in order to anticipate their trajectories and act effectively” [3].

To support people quickly make sense of big data, emerging technologies such as data mining, cloud computing and semantic web are being exploited to deal with huge volume of user generated information. In particular, the semantic web promises dramatic benefits by transforming unstructured and semi-structured documents into a “web of data”, which can be shared and reused across applications, enterprise, and community boundaries. It has great potential to support human in making sense of interconnected data by “connecting the dots” and “putting the pieces together”.

The problem of understanding and supporting sensemaking via technology remains challenging [14]. Against this background, we believe that a holistic view is needed to examine the different sensemaking models to identify essential

characteristics of sensemaking, and then consider whether the tool can support those characteristics. Therefore, in our study, we are interested in:

What are the shared characteristics of sensemaking that are revealed by sensemaking models?

How to conduct a user evaluation informed by these shared sensemaking characteristics in order to assess the impact of a tool on sensemaking activities?

Our proposed approach is driven by a priori understanding, through theories and models, of how human make sense of data and then apply that understanding to evaluate a chosen tool that support data exploration. The actual technical underpinning of the tool ceases to become the focus of the evaluation as it is the interaction with the user that would shed light for our understanding. In this paper, we use forum exploration as a case study of the wider data exploration problem. This paper is organized as follows: Section II provides related work on evaluation of sensemaking support tools; Section III describes sensemaking characteristics and measurements for evaluating those characteristics. An example of evaluation on sensemaking characteristics is given in Section IV.

2 Related Work

In order to help people understand the data more thoroughly and develop insight about the data, sensemaking tools usually are developed in the technical area of information retrieval, visual analytics and human computer interaction. Kittur et al. [6] presented a tool to assist users in organizing and understanding large collections of information using graph mining algorithms to. Maynard and Funk [7] made use of a variety of sophisticated natural language processing techniques to extract more meaningful and higher quality opinions from micro-blogs. Although many technology-driven tools are being built, their impact on sensemaking activities is not always obvious.

Some evaluation methods for sensemaking support tools have emerged but vary in terms of focus, measurements and analysis. For example, Scholtz [13] reported progress and challenges in evaluating tools for sensemaking with a list of metrics resulted from a workshop discussion for visual analysis environments. A model-driven formative evaluation under a sensemaking framework was presented by Qu and Furnas [11] in an evaluation of exploratory search, which examined interactive processes of representation construction and information seeking. Kang et al. [2] conducted an evaluation of a visual analytics system Jigsaw, which was employed in a small investigative sensemaking exercise and participants were observed against a sensemaking model. From their observations, a set of design implications were suggested to support a particular stage of sensemaking.

The above attempts have shown that sensemaking theories and models can make contributions to the design space of sensemaking support tools. However, most of those frameworks or models are presented as high level principles, and little guidance is provided for their application for the design and evaluation of sensemaking support tools.

3 Sensemaking characteristics

Sensemaking is an integral part of analysis. Sensemaking models have revealed various characteristics of the analytical processes of intelligence analysts. The purpose of identifying different characteristics of sensemaking from different disciplines is to bring together these different perspectives for a more holistic understanding of sensemaking. In our study, we have examined five models: Situation-gap-outcome [1], Cost Structure [12], Information Transformation [9], Data-frame Model [4] and Sensemaking Interaction [5]. From these models, we have identified four shared sensemaking characteristics in terms of: cognitive process, knowledge discovery, external representation and sensemaking experience. A summary of their supporting statements on these characteristics are shown in the Appendix. We believe that human factors in sensemaking should be included in any evaluation of a sensemaking support tool, and this can be achieved by obtaining feedback on whether the tool has any support for these four characteristics. To guide the analysis, we propose a list of questions which will help the design of evaluation for these characteristics (Table I).

TABLE I. GENERIC EVALUATION OBJECTIVES OF SENSEMAKING CHARACTERISTICS

Sensemaking Characteristics	Generic Evaluation Objectives
Cognitive Process	What sensemaking process (such as seeking, filtering, extracting etc.) does the tool support? Has the tool reduced any cost (time spent) on task to be performed?
Knowledge Discovery	What information transformation has the tool performed towards a knowledge product? Has the tool help user gain any knowledge?
External Representation	What representation has the tool provide for the user to complete the task? Has this representation aid the task to be performed?
Sensemaking Experience	What gap does the tool support the user in terms of constructing ideas, thoughts, emotions feelings or memories? Has the tool help user think deeper, more precisely or more broadly?

4 Evaluation of Forum Exploration Tool

Informed by sensemaking models, our evaluation approach involves three steps:

Step 1: Specify evaluation objectives for the tool in terms of sensemaking characteristics: cognitive process, knowledge discovery, external representation and sensemaking interaction.

Step 2: Design sensemaking scenario for user to initiate their sensemaking processes using the tool, including tasks and questions (data collection tool) to assess whether the tool has any impact on the sensemaking characteristics.

Step 3: Conduct quantitative and qualitative data analysis on the collecting data against identified objectives.

In the subsections below, we describe the experimental platform, Forum Exploration Tool (FET) with a selected Internet Forum, and how we evaluated the tool with the proposed approach.

4.1 Forum Exploration Tool and Dataset

FET was developed as one of Dicode¹ services to support the summarization of the content in Internet forums. Motivated by a Dicode use case, we selected a forum from StackExchange² which had over 1.5 million users across diverse topics from programming to cooking. Statistical analysis forum was chosen to simulate research scientists who used statistics in their work and were looking for tips from this forum. The output of FET consisted of 10 groups of topics, representing the most frequently discussed topic groups from this forum. Each group was presented as a topic cloud, with a title which showed the most discussed topic in that group. Other topics in the clouds were presented in different font sizes, proportional to the number of discussions related to that topic.

For example, Figure 1 shows four of the ten groups of topics extracted from statistical analysis forum, “Cross Validated”. The topic “Distribution” was extracted from 307 discussions and related topics are shown in this cloud, such as “normal”, “probability” and “sample”. By clicking a topic in the topic cloud (group title or words in the group), users can see all the related discussion threads from the group. By clicking on the title of the question, users can read the full detail of the questions and its answers.

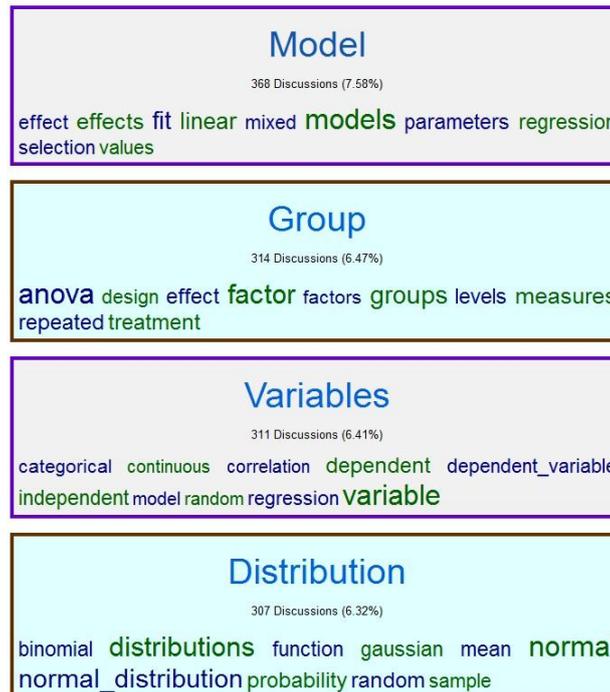


Figure 1 Example of topic clouds generated by the Forum Exploration Tool

¹ www.dicode-project.eu

² stackexchange.com

4.2 Specify Evaluation Objectives

From sensemaking point of view, FET could provide a quick overview of the topics of the selected forum discussions without reading into the details of the discussion. It also provided the related topics in one topic group, which enabled exploration of related topics within that group. To evaluate whether this tool could support user in making sense of the Internet Forum, we identified specific measurements and objectives against the sensemaking characteristics as shown in Table II.

TABLE II. FET EVALUATION OBJECTIVES AND MEASUREMENTS OF SENEMAKING CHARACTERISTICS

Sensemaking Characteristics	Evaluation Objectives [O] & Measurements [M] - FET
Cognitive Process	<p><i>[O] Has FET helped user gain a quick overview of the forum?</i></p> <p><i>[O] Has FET helped user find relationships among the discussions?</i></p> <p><i>[M]How many minutes did user spend on different tasks?</i></p> <p><i>[M] How difficult did user think for different tasks?</i></p>
Knowledge Discovery	<p><i>[O] Has FET help user discover knowledge among discussions?</i></p> <p><i>[M]How did user complete the task and why?</i></p> <p><i>[M]Is there any relationship between user background (knowledge and experience) and their process of completing the tasks?</i></p>
External Representation	<p><i>[O] Have the different functionalites of FET made any difference for completing the task?</i></p> <p><i>[M]Which functionalities did user use to complete the task?</i></p> <p><i>[M] Why did user choose one way over the other?</i></p>
Sensemaking Experience	<p><i>[O] Has FET enabled user think deeper in the process of exploration?</i></p> <p><i>[M] Was user confident using the tool for different task?</i></p> <p><i>[M] Did the tool provide any surprising result to the user?</i></p>

4.3 Design Sensemaking Scenario

A scenario was designed which included three tasks to initiate two sensemaking processes: “read & extract” and “search for relations” in information foraging [8]. This was given to the participants at the evaluation session.

Imaging that you are a member of a research team, which works on analysing a large volume of data to discover meaningful patterns (for example a bioinformatics research team analysing clinico-genomic data).

Your team conducts pre-processing of the data, setting the correct sample size for testing hypotheses, analyzing the data, and presenting the findings. The team has learnt that forums where people ask relevant questions about statistical analysis can provide helpful information, as it shows recent topics in this area and shares practical experience between people with different expertise. One such forum is “Cross Validated” where people conducting statistical analysis discuss issues they

have encountered. You have been asked by your team to check this forum and report back to them. You are required to prepare three presentation slides to the team using the templates provided. You have 60 minutes at maximum to do this task. The title and subtitles of the presentation slides have been pre-defined in the template shown in Figure 2.

<p style="text-align: center;">Overview of the Forum</p> <p><i>Identify 3 most discussed issues related to statistical analysis:</i></p> <p>No.1: No.2: No.3:</p> <p><i>Why and how did you identify these issues? (i.e. where you got the content and your reasons for selecting the three issues?)</i></p>
<p style="text-align: center;">Issues related to Normal Distribution</p> <p><i>Identify 3 most discussed issues related to Normal Distribution:</i></p> <p>No.1: No.2: No.3:</p> <p><i>Why and how did you identify these issues? (i.e. where you got the content and your reasons for selecting the three issues?)</i></p>
<p style="text-align: center;">Linking Sample Size and Hypothesis Testing</p> <p><i>Based on the discussions in this forum, please write down 3 statements that can relate "Sample size" with "Hypothesis testing".</i></p> <p>No.1: No.2: No.3:</p> <p><i>Why and how did you arrive at these statements? (i.e. where you got the ideas from?)</i></p>

Figure 2 Templates for the three tasks in FET evaluation

The first task (overview of the forum) and second task (issues related to normal distribution) were designed with the aim to initiate the sensemaking process: read & extract. This is because to complete the task, the participants have to read a collection of information and extract the issues from the reading. The hypothesis related to this task is:

[H1] The topic clouds have the potential to support "read and extract" process as the service has transformed raw information to a summarized format. If the topic clouds are trustworthy to the participants, they should get answers from the topic

clouds without checking the original discussions. This can save time spending on reading and extracting.

The third task (linking sample size and hypothesis testing) was designed with the aim to initiate the sensemaking process: search for relations. The hypothesis related to this task is:

[H2] The topic clouds have potential to support “search for relations” process. If the clouds are trustworthy to the participants, they should get answers from the topic clouds without checking the original discussions, which save the time spending on reading and extracting.

4.4 Data analysis

This study was conducted with ten participants from three European countries - UK, Spain and Greece. The outcome is summarised below.

Result I: Time spent on each task

During the study, we asked the participants to record the time they spent on each slide (task). The purpose was to gauge if participants spent more time on one task than the others. We analysed the data using a signed-rank test on two groups of pairs on one-sided hypothesis: task 1 < task 3 and task 2 < task 3. For those two groups of pairs, there were two nominal variables: task and participant, and one measurement variable (time of completion). The results showed that the participants did not spend significantly more time on task 3 than task 1 ($p=0.07$) or Task 2 ($p=0.05$).

Result II: User Confidence and Task Difficulty

Most of the participants (70% for task 1, 80% for task 2 and 80% for task 3) showed they were either very confident or confident in completing all three tasks and 40% the participants in statistical analysis forum claimed that task 3 is neutral or difficult (Figure 3). Participants who felt task 3 was either difficult or very difficult provided the following comments:

“It was a bit difficult to find statements for this task.”

“I did not found the related topics that I was looking for in the topic clouds.”

“Sorry, but after 15 minutes of searching, I can't say I found something to type.”

“There is no search function to help me complete this task.”

There was no obvious change in confidence level but there was a drop in difficulty level from task 2 to task 3 (90% to 60% of participants claimed the task is easy or very easy.) In contrast to the confidence level, the difficulty level (easy or very easy) in task 1 and task 3 claimed by the participants were almost the same (60%).

The subjective feedback of confident participants for task 1 and task 2 showed their thinking paths while interpreting the questions: “most discussed issues”: Participants who believed that the most discussed issues were shown in the topic clouds went to examine the topic clouds. Participants who believed the most discussed issues were “most viewed” or “most replied” questions went to examine the

lists of discussions. Participants who looked at both topic clouds and list of questions showed lower confidence level because they could not find consistent answers.

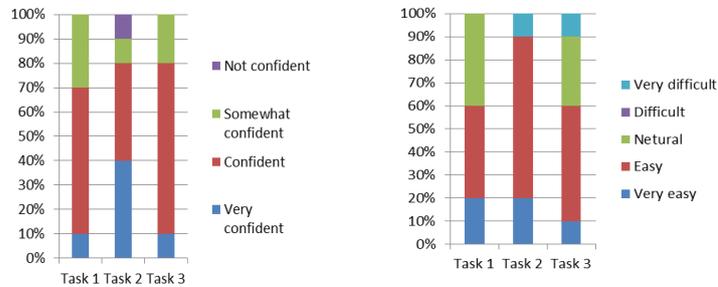


Figure 3 Confidence and difficulty on three different tasks

Result III: Topic clouds versus discussion list

Participants showed different opinions when performing different tasks. For task 1, half of the participants believed the list was more helpful while the other half believed the topic clouds were more useful. For task 2, more participants (60%) believed topic clouds were more helpful and among the rest of the participants, half (20%) believed the discussion list was more helpful and half (20%) believed both functions are helpful. For task 3, less than half participants (40%) believed topic clouds were more helpful, 30% of participants believed the list was more helpful and another 30% believed both functions were helpful.

Result IV: Strength and weakness of topic clouds

During the study, we collected feedback on the strengths and weaknesses of topic clouds completing their sensemaking tasks in order to inform future design.

Strengths

1. Visualizing the essence of the content
"Hierarchical structure"
"Better visualization of results"
"Clear layout, easy to see which topic is being used"
"A good number of related topics can be shown"
2. Organizing the topic in groups and easy for browsing the topics
"Quick and easy to use because of the topics are organized"
"It is easy and fast to access information"
"Reasonable clustering of the topics"
"Simple, clear and easy to locate main topics"
3. Easy for presenting simple summary
"Overall view of the forum, the importance of each term topic"
"In general, it provides easy information and a summary of the forums rather than reading through the thread discussions."

Weaknesses

1. No searching or sorting functions

“It doesn’t have search function”

“It lacks of any search and sort feature. It would be good to have a combination feature to select multiple topics and see the results. I think this would be useful in questions like the one at task 3.”

2. Not clear in relationships, No hierarchy (i.e. clouds within clouds)

“It would have helped if clicking the clouds within clouds would have provided a detailed level of clouds rather just giving the discussions they were in”

“Better to sort clouds in clouds in descending order for items with the same font size.”

“It could be more interesting to have another kind of representation as overlapped circles with common terms in intersections - When you click in a term you cannot look up over the list of discussions you get.”

3. Visual clues are not very clear: colouring, font, layout etc.

“When you click on a term you cannot search over the retrieved list of answers - A representation based on circles with a size depending on the relevance of the topic and intersections where common terms are located could be more graphical and representative”

4. No details about the topic

“Sometimes you don't know exactly what the topic is about just looking at the clouds”

4.5 Summary of Evaluation

To conclude, we revisited the objectives set before and summarized the result in Table III.

TABLE III. FET EVALUATION CONCLUSION OF SENSEMAKING CHARACTERISTICS

Sensemaking Characteristics	Evaluation Objectives [O] & Conclusion [C] - FET
Cognitive Process	<p><i>[O1] Has FET helped user gain a quick overview of the forum?</i></p> <p><i>[C1]The FET has helped users gain a quick overview because more than half of the participants completed task 1 and task 2 using the topic clouds, which saved them time reading the original discussions.</i></p> <p><i>[O2] Has FET helped user find relationships among the discussions?</i></p> <p><i>[C2]The FET has not helped users find relationships among the discussions because users felt the task of finding relationships (task 3) difficult and no existing functionalities were helpful.</i></p>
Knowledge Discovery	<p><i>[O] Has FET helped user discover knowledge among discussions?</i></p> <p><i>[C]The feedback from users showed their thinking path from interpreting a question to which functionalities to use for seeking an answer. Unfortunately,</i></p>

Sensemaking Characteristics	Evaluation Objectives [O] & Conclusion [C] - FET
	<i>we could not build any relationship between user profile and their thinking path because of the small number of the participants.</i>
External Representation	<p><i>[O] Have the different functionalities of FET interface made any difference for completing the task?</i></p> <p><i>[C] Two functionalities (topic clouds and list of questions) of FET did not make any difference for users completing the task. Users provided their feedback on improving the two functionalities, such as visual cues, multiple term query, multiple level of topics etc.</i></p>
Sensemaking Experience	<p><i>[O] Has FET enabled user think deeper in the process of exploration?</i></p> <p><i>[C] Users were confident with completing all tasks although they felt task 3 was more difficult. However, their thinking paths to solve task 3 did not show that they think deeper enough to get a satisfied answer. In addition to that, they did not spent more time on task 3 and only a small percentage of participants used both functionalities to complete the task.</i></p>

5 Discussions and Future Work

This paper has presented a preliminary work on investigating human factors in forum exploration from a sensemaking perspective. It is a model-driven approach drawing from literature of cognitive science in order to identify essential activities of people making sense of data. From examining the five different models, four characteristics of sensemaking (cognitive process, knowledge discovery, external representation and user experience) were identified. We consider those four characteristics as human factors because they are important human activities and behaviour during sensemaking. This approach has provided two benefits. Firstly, the identified characteristics of sensemaking provided a generic framework for setting criteria in the evaluation of a sensemaking support tool. Secondly, the experimental study guided by this framework gave us an understanding of how the tool was being used for the sensemaking tasks.

However, the four characteristics are just a starting point to investigate sensemaking models for evaluating sensemaking support tools. It would be interesting to see whether different characteristics can be derived from same set of models by different researchers or more characteristics can be derived from more models. The challenge also fell on the translation from high-level criteria to low-level measurements for data collection and analysis. This translation involved understanding the models and selecting appropriate measurements (see Table I). From this study, we have learned that understanding the technical features of the tool is also necessary because we need a realistic expectation on the value added by the tool in order to design appropriate sensemaking tasks. Furthermore, considerable effort was spent in designing the scenario to trigger sensemaking activities which were feasible for a controlled lab experiment. . It would be worth exploring whether the scenario and tasks can be transferred for evaluating other sensemaking tools.

In the near future, we plan to continue this work towards a general set of questions for evaluating sensemaking tools through conducting more evaluation studies with more users involved.

Acknowledgements

The research leading to these results has received funding from the European Union Seventh Framework Programme (FP7/2007-2013) under grant agreement no ICT 257184 (DICODE project). Thanks go to users participated in the evaluation study for their time and valuable feedback.

6 References

1. Dervin, B. (1983). An Overview of Sensemaking Research: Concepts, Methods, and Results to Date,"Int'l Communication Assoc., 1983;
2. Kang, Y.-a., Görg, C. and Stasko, J. (2009). Evaluating Visual Analytics Systems for Investigative Analysis: Deriving Design Principles from a Case Study. in IEEE Symposium on Visual Analytics Science and Technology.
3. Klein, G., Moon, B. and Hoffman, R. R. (2006). Making Sense of Sensemaking 1: Alternative Perspectives. IEEE Intelligent Systems, 21(4), pp. 70-73.
4. Klein, G., Moon, B. and Hoffman, R. R. (2006). Making Sense of Sensemaking 2: A Macrocognitive Model. IEEE Intelligent Systems, 21 (5), 2006, 88-92.
5. Kirsh, D. (2009). Interaction, External Representations and Sense Making. In N. A. Taatgen & H. van Rijn (Eds.), Proceedings of the 31st Annual Conference of the Cognitive Science Society (pp. 1103-1108). Austin, TX: Cognitive Science Society.
6. Kittur, A., Chau, D.H., Faloutsos, C. and Hong, J. I. (2009). Supporting Ad Hoc Sensemaking: Integrating Cognitive, HCI, and Data Mining Approaches, In: Sensemaking Workshop at CHI, Boston, M.A
7. Maynard, D., and Funk, A. (2011). Automatic detection of political opinions in tweets. In: Proc. of MSM 2011: Making Sense of Microposts Workshop at 8th Extended Semantic Web Conference, Heraklion, Greece, May 2011.
8. Pirolli, P. and Card, S.(1999). Information Foraging. Psychological Review, 106, pp. 643-675.
9. Pirolli, P. and Card, S.(2005). The sensemaking process and leverage points for analyst technology as identified through cognitive task analysis. Proceedings of International Conference on Intelligence Analysis.
10. Pirolli, P. and Russell, D. (2011). Introduction to this Special Issue on Sensemaking, Human-Computer Interaction, 26 (1-2), pp.1-8
11. Qu, Y. and Furnas, G. (2008). Model-driven formative evaluation of exploratory search: A study under a sensemaking framework. Information Processing & Management, 44(2), pp. 534-555.
12. Russell, D.M., Stefik, M.J., Pirolli, P., and Card, S.K. (1993). The cost structure of sensemaking. Proc. SIGCHI, ACM Press New York, NY, USA (1993), 269-276.
13. Scholtz, J. (2008) Progress and Challenges in Evaluating Tools for Sensemaking. ACM Computer Human Information (CHI) conference Workshop on Sensemaking in Florence, Italy, April 6, 2008.
14. Whittaker, S. Making sense of sensemaking. In Erickson, T. and McDonald D.W. (Eds.), HCI remixed: Reflections on works that have influenced the HCI community. Boston: MIT Press (2008), 173-178.

APPENDIX: SENSEMAKING CHARACTERISTIC IN SENSEMAKING MODELS

Sensemaking Models	Sensemaking Characteristics			
	<i>Cognitive Processes</i>	<i>Knowledge Discovery</i>	<i>External Representation</i>	<i>Sensemaking Experience</i>
Situation-gap-outcome [Dervin, 1983]		The person facing this gap constructs bridges consisting of ideas, thoughts, emotions, feelings and memories.		Sense-making occurs when a person embedded in a particular context and moving through time-space, experiences a “gap” in reality.
Cost structure [Russell et al., 1993]	Sensemaking is modeled as cyclic processes of searching for external representations and encoding information into these representations to reduce the cost of tasks to be performed.	It involves not only finding information, but also requires learning about new domains, solving ill-structured problems, acquiring situation awareness, and participating in social exchanges of knowledge.	Sensemaking is the process of searching for a representation and encoding data in that representation to answer task-specific questions.	
Information Transformation [Pirolli and Card, 2005]	This process consists of two loops of activities: a foraging loop that involves seeking, filtering, and extracting information into schemas; and a sensemaking loop that involves iterative development of a mental model from the schemas that best fit the evidence.	Sensemaking is a process of transformation of information into a knowledge product. This processing can be driven by bottom-up processes (from data to theory) or top-down (from theory to data).	Such tasks consist of information gathering, representation of the information in a schema that aids analysis, the development of insight through the manipulation of this representation, and the creation of some knowledge product or direct action based on the insight.	
Data/Frame Theory [Klein et al., 2006]	Sensemaking is a motivated, continuous effort to understand connections (which can be among people, places, and events) in order to anticipate their trajectories and act effectively.	The phenomenon of sensemaking ties also to the notion that frames are chunks of knowledge abstracted away from computational details—symbolic descriptions that are taken off the shelf and used to perceive things, and thereby constitute understanding.		
Sensemaking Interaction [Kirsh, 2009]			Sensemaking interaction is referred to the interaction concerned with external representation and it is our reliance on tools, representations and techniques for solving problems and externalizing thought	Through (sensemaking) interaction it is easier to process more efficiently (i.e. with better speed-accuracy), more deeply, more precisely, and often more broadly than by working inside the head alone.