Human Factors Evaluation of Shorthand Notations in Sharik

Shadi Ghajar-Khosravi
Peter Kwantes

DRDC – Toronto Research Centre

Defence Research and Development Canada

Abstract

Sharik is a collaboration support tool that allows users to express and share information with team members. The key capability of the tool is enabling analysts to share their insights, discoveries and hypotheses as propositions. We treat the proposition as the basic unit of information in an analysis. To allow users to enter propositions into Sharik in a fast and efficient manner, we have devised a syntax that users can use to enter information at a command line. The purpose of this reference document is to report the results of a user study which was conducted to measure the extent to which users can gain proficiency in entering propositions using the syntax. The feedback collected indicated that although it took the participants some time to learn the concept and structure of propositions, they managed to learn and use the syntax in a fast and efficient manner.

Significance to defence and security

Analysts are continuously dealing with larger amounts of information to review, annotate, analyse, and extract intelligence from. A wide range of solutions is proposed for tackling this information overload issue, including various analytic tools supporting analysts' analyses. On a different spectrum, another solution would be facilitating collaboration, data sharing, and sensemaking among teams of analysts working together on a common mission with the ultimate goal of improving efficiency and reducing redundancy in intelligence analysis activities. Hence, Sharik with its various features (including command line) was designed to support collaborative sensemaking among intelligence analysts.

This work is being carried out at Defence Research and Development Canada (DRDC) under project 05da: JICAC (Joint Intelligence Collection and Analysis Capability) within the Joint Force Development (JFD) S&T portfolio. The JICAC project is being run as a collaboration between DRDC's Valcartier and Toronto Research Centres, and aims to provide the Canadian Armed Forces (CAF) with tools, techniques, and advice for procurement to reduce the amount of time all-source intelligence analysts must search for information, and maximise the amount of time they can spend doing analysis.

Résumé

Sharik est un outil de soutien à la collaboration qui permet aux utilisateurs de s'exprimer et de partager de l'information avec les membres de leur équipe. Cet outil a comme principal atout de permettre aux analystes de partager leurs connaissances, leurs découvertes et leurs hypothèses sous forme de propositions. Dans une analyse, nous traitons la proposition comme unité de base de l'information. Afin de permettre aux utilisateurs d'entrer les propositions dans Sharik de manière rapide et efficace, nous avons conçu une syntaxe que les utilisateurs peuvent employer pour saisir l'information dans une ligne de commande. Ce document de référence vise à rendre compte des résultats d'une étude sur les utilisateurs, qui a été menée pour déterminer dans quelle mesure les utilisateurs maîtrisent l'entrée des propositions au moyen de la syntaxe. Selon les rétroactions recueillies, nous savons que même si les participants ont mis du temps à apprendre le concept et la structure des propositions, ils ont tout de même appris la syntaxe et s'en servent de manière rapide et efficace.

Importance pour la défense et la sécurité

Le volume d'information que les analystes doivent examiner, annoter et analyser afin d'extraire des renseignements augmente sans cesse. Par conséquent, il existe un vaste éventail de solutions visant à composer avec cette surabondance d'information, dont divers outils d'analyse. Or, il existe également une solution différente qui facilite la collaboration, le partage de données et la recherche de sens au sein d'équipes d'analystes travaillant ensemble dans le cadre d'une mission commune et dont l'objectif ultime est d'améliorer l'efficacité et de réduire la redondance des activités d'analyse du renseignement. Ainsi, Sharik et ses diverses fonctions (y compris la ligne de commande) ont été conçus pour appuyer la recherche de sens collaborative parmi les analystes du renseignement.

Les travaux se déroulent à Recherche et Développement pour la défense Canada (RDDC) dans le cadre du projet 05da: JICAC (Projet interarmées de collecte de renseignements et de la capacité d'analyse) au sein du portefeuille de S & T du Développement de la force interarmées (DFI). Les centres de recherche de RDDC Valcartier et Toronto collaborent à ce projet qui vise à fournir aux Forces armées canadiennes (FAC) des outils, des techniques et des conseils en matière d'approvisionnement afin de réduire le temps que les analystes du renseignement toutes sources passent à chercher des renseignements et de maximiser le temps qu'ils consacrent à l'analyse.

Table of contents

List of figures

List of tables

Acknowledgements

1   Introduction

Sharik is a browser-based software platform to support analysis, or sensemaking, done in teams. The tool discussed in this report is called Sharik (SHAring Resources, Information, and Knowledge), a web-based tool aimed at supporting analysis and collaborative sensemaking among all-source intelligence analysts in either distributed or co-located intelligence analysis locales (e.g., All-Source Intelligence Centres, or ASICs).

The key capability of the tool is enabling analysts to share their insights, discoveries and hypotheses as propositions. We consider a proposition as the basic unit of information in an analysis. In Sharik, a proposition is a pair of concepts connected by a linking phrase to form a statement that can be evaluated as being either true or false. For example, "*John loves Mary*" is a proposition, with the concepts John and Mary being linked by the term "love" to form a statement that is either true or not true. Analysts are meant to enter their propositions into a database so they can be shared across users. In our conversations with analysts, we have been advised that any new tools to support analysis should do so with the least number of keystrokes and mouse clicks possible.

To allow users to enter propositions into Sharik in a fast and efficient manner, we have implemented a command line feature to avoid the need for users to exploit point-and-click on text boxes. We reasoned that a command line, once mastered, would be an efficient and preferable way to enter information into the system compared to a point-and-click interface. We have devised a syntax that users can employ while entering propositions of different types at the command line.

Propositions in Sharik take three different forms. Propositions that represent things we know to be true without any supporting evidence are called, *facts*. *Extracts* are facts we know to be true because we have supporting evidence. Finally, *conjectures* are propositions that may or may not be true. Propositions can be unidirectional or bidirectional. For example, *John loves Mary* is unidirectional, because it does not follow that Mary also loves John. John is married to Mary is bidirectional because, if John is married to Mary, then Mary is also married to John. Once in the system, Sharik integrates the propositions supplied by all of the users to create a concept map (CMap) to display the information contributed by all analysts in a common visualization. See Figure 1.

The purpose of this reference document is to report the results of a user study that was conducted in Fall 2016 to evaluate how quickly users can enter propositions using the syntax at the command line with practice.

Figure 1: *An example CMap automatically created by Sharik.*

2 Participants

Five participants took part in the study. There were recruited from DRDC – Toronto Research Center. Four participants were male.

3   Method

This human factors study was designed to measure the extent to which users can gain proficiency with the syntax with training. The study took place at DRDC – Toronto Research Center and involved participants entering their responses in Windows command prompt on a laptop computer. The study consisted of three phases:

Phase 1: In the first phase, participants were given training as to how to write propositions using the syntax. Specifically, participants were given instructions as to what a proposition is, and how they are expressed in the Sharik syntax. The training was paper-based and prepared in a three-page training sheet format. The training sheet had two primary components. In the first, propositions and their categories were defined. In the second component, the shorthand syntax is explained along with multiple examples under each variation. The training sheet is provided in Appendix A and the example tables in Appendix B.

Phase 2: In the second phase, participants practised writing sentences in the syntax using a specially-written experimental program. The program was coded in Java and executed at a Windows command prompt. See Figure 2 for a screenshot of the program. Participants started with practise trials which only consisted of three propositions and provided an opportunity to familiarize themselves with the tool's prompts and ask any questions on the steps.

Once the trial was completed, the experiment started. The tool displayed 72 sentences[1] and asked the participants to rewrite the sentences as propositions using the syntax for which they received training in Phase 1. The 72 propositions were randomly ordered for each participants and divided into six blocks and participants had an optional break after each block, i.e., after every 12 sentences/propositions. For each sentence, we measured the time from its appearance on screen to when the 'enter' key was pressed after it had been written as a proposition.

In the final block, they were instructed to copy the 12 sentences verbatim rather in shorthand format. The point of the last block of trials was to obtain completion times for the sentences so that we can measure the time cost in transcribing the sentences into propositions. Overall, six blocks of 12 trials (i.e., 72 sentences) were completed by the end of each study.

Participants were allowed to ask any questions about the content of the training sheet as well as the experiment's process before the actual experiment started. However, they were advised not to ask any questions during the experiment. They were allowed to refer back to the training sheet and example tables during the experiment.

Phase 3. After completing the second phase, participants were asked four open-ended follow-up questions about their experience with the syntax (Appendix C).

---

[1] In the study's protocol, the number of sentences was specified 96 in total. However, we cut two blocks based on the results of a pilot study which took longer than one hour. Cutting the two blocks enabled us to finish all the studies in less than hour.

Figure 2: *Experiment tool screenshot.*

4   Results

The study took approximately an hour for all the participants. All five participants completed the study.

4.1   Task completion accuracy

Participants were marked based on their answers. Each proposition was graded on a 5-point scale based on the following criteria:

- Correctly assigning the proposition's direction: 1 point

- Correctly connecting two entities with an appropriate linking phrase: 2 points

- Correctly differentiating facts from hypothesis: 1 point

- Correctly assigning properties: 2 points

If a proposition was not a hypothesis or did not have any properties, they still received the full 3 points for them. Table 2 shows the total grade for each participant calculated as the percentage of possible points earned.

Participants lost marks mostly on:

- Identifying the direction of propositions (i.e., directional vs. bidirectional)

- Identifying the appropriate linking phrase and second entity. Similarly, deciding on what goes into property section and what goes into second entity. We envision that in a real-world investigation, analysts would have a deeper understanding about the key entities of an investigation and the choices of entities and properties would be more natural and clear.

- Only one of the participants showed difficulty and lost marks on the property labels/values syntax. The rest of the participants were almost perfect in particular in the final three shorthand writing blocks. Although some participants reported confusions regarding properties, they made very few mistakes.

Accuracy was high over-all, with an average mark over five participants of 79% and the lowest performance being 76%. See Table 1.

Table 1: The marks received by participants 1 to 5.

	Participant 1	**Participant 2**	**Participant 3**	**Participant 4**	**Participant 5**	**Avg.**
Total Mark	81%	79%	76%	79%	80%	79%

DRDC-RDDC-2016-D085

4.2 Task completion time

Participants were timed on all 72 sentences that they entered. Generally speaking, all participants performed faster as they moved from one block to the next. The modesty of the change could be due to the syntax being easy leaving little room for improvement or difficult to make any improvements in the time provided. The feedback we report in the next section and the high task completion success marks (see previous section) point to the former case.

Figure 3. *Average task completion time (secs) across blocks.*

As expected, all five participants performed faster in the verbatim block (Figure 3). However, as our participants noted in the subsequent discussion, the time difference is mostly due to the translation processes involved in blocks 1 to 5. Deciding on what portions of sentences should go to which segments of the propositions was found to be a more time consuming aspect of the task than putting the proposition into shorthand format.

4.3 Closing questions

After the participants had completed the proposition task they were asked five questions about their impression of the syntax. The following tables summarize the five participants' responses to the four closing questions.

Table 2: *Summary of responses to "What is your overall impression of this syntax?".*

Participant ID	Overall Impression
1	Good impression overall
2	Easy to pick up
3	Found it relatively easy to learn – had to rely on training sheet less and less throughout the study
4	Easy to learn and intuitive
5	I did not particularly like it

Table 3: *Summary of responses to "What aspects of the syntax did you like the best?".*

Participant ID	Liked Best
1	- I liked the structure. It was clear how to translate sentences into the syntax.
2	- It was good for summarizing concepts quickly and pretty straightforward to add properties
3	- I liked the flexibility of the syntax
4	- ToDate/FromDate was clear - I liked the question mark to indicate hypothesis because it was clear and good for conciseness - There were few rules to remember
5	- I thought the properties were clear and simple

Table 4: *Summary of responses to "What aspects of the syntax did you like the best?".*

Participant ID	Liked least
1	- Deciding what should be a property and what the label should be
2	- When there are three nouns, hard to know what to include as part of a concept vs. property - I wasn't sure whether it was necessary to switch from active to passive voice or vice versa - Some adverbs weren't included, such as 'initially' - Didn't include articles (i.e. 'the') most of the time – might change meaning of proposition
3	- I was confused about when to include 'at' and 'in', - When properties were about the concept, not the linking phrase, felt like she should put a colon after the second concept (e.g.. Toronto District School board: Programs) - I didn't know whether to include tenses or not
4	- Deciding on where to draw the line from linking phrase to concept two. For example, in 'was visited by' I was not sure where to cut off linking phrase or whether or not to include 'by'
5	- I found the directionality confusing, thought that should be taken care of automatically by the computer

Table 5: Summary of responses to "Do you have any final comments or suggestions?".

Participant No.	Overall Impression
1	There were not that many hypothesis examples in the study. So, I often forgot to include the question mark. I was unsure whether to include 'is' and 'the' at first, but then figured it out
2	None
3	Be more specific about when to use a property and make it more clear how to record multiple dates
4	None
5	None

5   Discussion and conclusion

Although, the participants of this study were trained on the definition and applications of propositions, it is assumed that in the operational setting, new users of Sharik will already have sufficient training on the propositions before using it. Hence, we speculate that the confusions regarding what makes the first and second entity, and the make-up of the linking phrases would be minimal when operational. Deciding on what portions of sentences should go to which segments of the propositions was found by the participants as the most time consuming aspect of the task rather than putting the propositions into shorthand format.

Given the shallow slope of the trend line in Figure 3, the fact that all participants had high accuracy, and the overall positive verbal feedback, we conclude that the actual syntax is easy to learn. However, moving forward, we should pay special attention to how training is conducted to ensure that transforming sentences into propositions is as clear as possible.

The current syntax allows for both passive and active voice propositions. However, the syntax might become more easy to use if we add a new rule to the syntax that limits propositions to only active or passive voice format.

Based on the data collected in this study, we will update the syntax along with the training component and continue developing and coding the shorthand feature into Sharik.

Annex A   Training sheet

A.1   Propositions

A.1.1   What is a proposition?

A proposition is a statement **linking 2 concepts** in a way that is expressed as a judgment that could be true or false. In other words, propositions are expressed as **a triple** – two concepts connected by a linking phrase.

For example, **John works at DRDC** is a proposition consisting of 2 concepts and a linking phrase:

A.1.2   What is a concept?

A concept could be **anything** depending on the context of an investigation. Concepts are the **key entities** of interest (e.g., people, place, organizations, or even a quality) in an intelligence analysis investigation. They are mostly **nouns** but could be expressed in other word formats as well.

A.1.3   Fact or hypothesis?

Propositions could be factual statements where the analyst entering the proposition believes the statement is a fact. The proposition **John works at DRDC** is **a fact** or true statement lacking any doubts or uncertainty.

However, the proposition **John may live in Toronto** is not a fact. John may or may not live in Toronto. The statement is a **hypothesis** or guess and has uncertainty due to having the word "**May**" in it. Other sample words making a proposition a hypothesis would include: probably, likely, might, etc.

A.1.4   Unidirectional or bidirectional?

Propositions could be unidirectional or bidirectional. For example, **John is married to Mary** is **bidirectional** because we could also say **Mary is married to John**. However, **John loves Mary** is unidirectional. We cannot necessarily conclude that "Mary loves John".

A.1.5 Properties

Propositions could also have properties. Properties provide additional information about the link between the two concepts. For example **John married Jane in Toronto.** In this case, the main statement is that John married Jane. That the wedding happened in Toronto is a property of the wedding.

A.2 Shorthand syntax

We have devised **a syntax** for entering propositions into the online tool Sharik in **shorthand**.

Basic: Symbols **"-"** and **">"** are used to separate the two concepts from the linking phrases. For unidirectional propositions **"-"** and **">"** are used while for bidirectional propositions two instances of **"-"**.

The Taliban will carry out new large scale attacks
The Taliban **–** will carry out **>** New Large scale attacks

The Taliban leaders met with Al-Qaida leaders
The Taliban leaders **–** met with **–** AlQaida leaders

Hypotheses? A hypothesis is identified by a **"?"** as the first character. A proposition without a **"?"** character is a fact.

The Taliban **probably** accepted the Doha office in 2013.
? The Taliban **–**accepted**>** Doha office

NOTE: When a hypothesis is entered in Sharik, the **uncertainty term is dropped** from the proposition (e.g., **"probably"** does not appear in the final syntax above). The "?" symbol conveys that there is uncertainty in the statement.

Properties: Properties are entered right after the proposition separated by a **","**. To enter a property the label/name of the property is entered first followed by the value of the property. The name and value of the properties are separated by **":"**:

John married Jane in Toronto
John **–** married **–** Jane**,** in:Toronto

In the example above, a comma is entered right after "Jane" to separate the property from the actual proposition. The context/name of the property is **"in"** and the value of the property is **"*Toronto*"** separated with a **":"** from the context.

in : Toronto

Property label   Separating colon   Property value

Multiple properties: A proposition can have multiple properties. A property is separated by a **","** from other properties.

John married Jane in Toronto during summer
John **–** married **–** Jane**,** in**:**Toronto**,** during**:**summer

---

Reserved <u>Date</u> property: We treat dates as a special type of property. Hence, we have three reserved words for date properties. They are, **date** (for a <u>particular occasion</u>), **fromDate** (indicating a <u>start date</u>), and **toDate** (indicating an <u>end date</u>).

The Date properties expect the date written in the formats, **YYYYMMDD**, **YYYYMM**, or **YYYY**

Suicide bombers assaulted police academy in Kabul on July 5th 2015.
Suicide bombers **–** assaulted **>** Police academy**,** in**:** Kabul**, Date:** 20150705

Molla Omar studied medicine at university of Doha, from 1970 to 1976.
MollaOmar **–** studied**>** medicine**,** at**:** university of Doha**, fromDate:**1970**, toDate:**1976

NOTE: Except for date properties (for which you should follow the standard labels above), you are free to choose **any labels** as for the property labels.

Repeated properties: If there are multiple occasions of a same property, the property labels should be numbered.

Molla Omar met with Osama Bin-Laden three times on July 5th 1980, June 21st 1982, and October 1998.
Molla Omar **–** met with **–** Osama Bin Laden**,** frequency**:** 3 times**, Date1:**19800705**, Date2:** 19820621**, Date3:**199810

NOTE: Space before or after **punctuations** is not mandatory. All these instances are correct:

John-works at>DRDC
John – works at> DRDC

John – works at > DRDC

_____

NOTE: There could be multiple correct shorthands for each proposition. The "Correct Answers" shown during the experiment are only one of the possible correct answers.

Sentence	Proposition	TYPE
Toronto is in Canada	Toronto -is in> Canada	Basic, unidirectional
John is married to Jane	John -is married to- Jane	Basic, bidirectional
Fred might be a doctor	? Fred -is a> doctor	Hypothesis, unidirectional
It is possible that Tony Stark is Iron Man	? Tony Stark -is- Iron Man	Hypothesis, bidirectional
New large scale attacks might target diplomatic assets	? New large scale attacks -- will target > diplomatic assets	Hypothesis, unidirectional
John married Jane in Toronto during summer	John -- married -- Jane, in:Toronto, when:summer	Multiple properties
Taliban negotiated with Afghans over peace talks in Qatar.	Taliban -- negotiated with -- Afghans, over: peace talks, in: Qatar	Multiple properties
US official offered a Doha office to the Taliban at May 2013 negotiations.	US officials -- offered> a Doha office, to: the Taliban, etc:2013 negotiations, Date:201305	Multiple properties, Date
Molla Omar studied medicine at university of Doha, from 1970 to 1976.	MollaOmar -- studied> medicine, etc university of Doha, fromDate:1970, toDate:1976	Multiple properties, fromDate, toDate
Molla Omar met with Osama Bin-Laden three times on July 5th 1980, June 21st 1982, and October 1998.	Molla Omar -- met with -- Osama Bin Laden, frequency: 3 times, Date1:19800705, Date2: 19820621, Date3:199810	Multiple properties, repeated properties

Annex C   Closing interview questions

1. What is your overall impression of this syntax?

2. What aspects of the syntax did you like the best?

3. What aspects of the syntax did you like the least?

4. Do you have any final comments or suggestions?

DRDC-RDDC-2016-D085

List of symbols/abbreviations/acronyms/initialisms

CAF	Canadian Armed Forces
CCIR	Commander's Critical Information Requirements
CCIRM	Collection Coordination and Intelligence Requirements Management
Cmap	Concept Map
DRDC	Defence Research and Development Canada
HUMINT	Human Intelligence
IC	Intelligence Cycle
ICP	Intelligence Collection Plan
IR	Information Requirement
JICAC	Joint Intelligence Collection and Analysis Capability
PIR	Priority Intelligence Requirement
OSINT	Open Source Intelligence
RFI	Request for Information
Sharik	SHAring Resources, Information, and Knowledge
SIGINT	Signal Intelligence
SME	Subject Matter Expert

13. ABSTRACT (A brief and factual summary of the document. It may also appear elsewhere in the body of the document itself. It is highly desirable that the abstract of classified documents be unclassified. Each paragraph of the abstract shall begin with an indication of the security classification of the information in the paragraph (unless the document itself is unclassified) represented as (S), (C), (R), or (U). It is not necessary to include here abstracts in both official languages unless the text is bilingual.)

Sharik is a collaboration support tool that allows users to express and share information with team members. The key capability of the tool is enabling analysts to share their insights, discoveries and hypotheses as propositions. We treat the proposition as the basic unit of information in an analysis. To allow users to enter propositions into Sharik in a fast and efficient manner, we have devised a syntax that users can use to enter information at a command line. The purpose of this reference document is to report the results of a user study which was conducted to measure the extent to which users can gain proficiency in entering propositions using the syntax. The feedback collected indicated that although it took the participants some time to learn the concept and structure of propositions, they managed to learn and use the syntax in a fast and efficient manner.

Sharik est un outil de soutien à la collaboration qui permet aux utilisateurs de s'exprimer et de partager de l'information avec les membres de leur équipe. Cet outil a comme principal atout de permettre aux analystes de partager leurs connaissances, leurs découvertes et leurs hypothèses sous forme de propositions. Dans une analyse, nous traitons la proposition comme unité de base de l'information. Afin de permettre aux utilisateurs d'entrer les propositions dans Sharik de manière rapide et efficace, nous avons conçu une syntaxe que les utilisateurs peuvent employer pour saisir l'information dans une ligne de commande. Ce document de référence vise à rendre compte des résultats d'une étude sur les utilisateurs, qui a été menée pour déterminer dans quelle mesure les utilisateurs maîtrisent l'entrée des propositions au moyen de la syntaxe. Selon les rétroactions recueillies, nous savons que même si les participants ont mis du temps à apprendre le concept et la structure des propositions, ils ont tout de même appris la syntaxe et s'en servent de manière rapide et efficace.

14. KEYWORDS, DESCRIPTORS or IDENTIFIERS (Technically meaningful terms or short phrases that characterize a document and could be helpful in cataloguing the document. They should be selected so that no security classification is required. Identifiers, such as equipment model designation, trade name, military project code name, geographic location may also be included. If possible keywords should be selected from a published thesaurus, e.g., Thesaurus of Engineering and Scientific Terms (TEST) and that thesaurus identified. If it is not possible to select indexing terms which are Unclassified, the classification of each should be indicated as with the title.)

Human Factors Evaluation; User Interface Design