Usefulness of Information for Goal Achievement

Laurence Cholvy, Célia da Costa Pereira

▶ **To cite this version:**

HAL Id: hal-02279935

https://hal.archives-ouvertes.fr/hal-02279935v2

Submitted on 7 Sep 2019

Usefulness of Information for Goal Achievement

Laurence Cholvy[1] and Célia da Costa Pereira[2] [*]

[1] ONERA, Toulouse
Laurence.Cholvy@onera.fr
[2] Université Côte d'Azur, CNRS, I3S Lab.
Celia.DA-COSTA-PEREIRA@univ-cotedazur.fr

Abstract. This paper focuses on modelling information usefulness. More precisely, it aims at characterizing how useful a piece of information is for a cognitive agent which has some beliefs and goals. The paper presents three different approaches. We take Information Retrieval as a particular application domain and we compare some existing measures with the usefulness measure introduced in the paper.

1 Introduction

Usefulness is a ubiquitous notion. For instance, in Data Mining, evaluating the interest of the extracted knowledge is necessary [9]; in Natural Language Processing, identifying useful terminology [19] is a prerequisite to any analysis. In Social Science, studying how people achieve effective conversational communication in common social situations is needed. Grice [10], introduced the maxim of quantity which emphasizes the fact that a speaker contribution must be as informative as required for the current purposes of the exchange, but not more informative. In the database domain, taking the goals and the preferences of the user who asks a query is necessary for generating cooperative answers [12]. In Information Retrieval (IR), the aim is to take into account a query expressed by a user and provide documents which best suit the user need i.e., which are the most useful ones. Initially, the *topical relevance* approach considered that relevant documents are those whose topics best match the topics of the user query [11]. This led to the *aboutness* measure. Then, other dimensions have been considered: *coverage*, which measures how strongly the user interests are included in a document [13]; *appropriateness*, which measures how suitable a document is with respect to the user interests [8]; and *novelty*, which measures how novel is the document with respect to what the system has already proposed to the user [4]. However, the user who asks a query is a cognitive agent [14,7]: he/she has some goals to achieve and he/she has some beliefs about the world. Moreover, these beliefs are generally incomplete and the user asks queries to the system in order to get new information which will help him/her achieve his/her goals.

In the present work, we consider a general framework in which there are two cognitive agents: one is the user who has some beliefs and some goals modelled as propositional formulas; the second is the system. The latter has some beliefs about the user's

beliefs and goals. Its goal is to provide the user with information which is the most useful for him/her to achieve his/her own goals. This framework is general enough to model the paradigm of cooperative exchanges with a system (a speaker, a database, the search engine) who answers the query expressed by the user (the listener, the database user, the web user...) in which the system has to provide the most useful information to the user. Defining the concept of information usefulness in such a context is the main aim of this paper. More precisely, we take the system point of view and try to characterize how useful a piece of information can be for the user.

The paper is organised as follows. In Section 2 we give some preliminaries and state our working hypotheses. In Sections 3, 4, 5 we propose three different definitions of information usefulness, respectively called binary, ordinal and numerical. In Section 6, we consider the particular case of Information Retrieval and compares some measures defined there with ours. Some concluding remarks are given in Section 7.

2   Preliminaries

We consider a propositional language L of which a subset, L_G, is the language used to represent the goals. We consider an agent a with a goal set G_a, which is a finite set of positive literals from L_G. For example, *finish the state of the art of my article, prepare for Monday's class*. Moreover, agent a has a belief base B_a composed of two subsets B_a^m and B_a^g. B_a^m is the set of formulas from $L \setminus L_G$ which represents a's beliefs. For example, *I know modal logic* and *I know the Python language*. B_a^g contains as many formulas $l_g^1 \wedge \ldots \wedge l_g^{m_g} \rightarrow g$, where each l_g^i is a positive literal of $L \setminus L_G$, as there are $g \in G_a$. Such formulas represent the beliefs of a about what is needed to achieve its goals. For example, *to finish the state of the art (g) I need knowledge about modal logic (p) and BDI agents (q)* (i.e., $p \wedge q \rightarrow g$). The conjunction $l_g^1 \wedge .. \wedge l_g^{m_g}$ is called *premise* of g and it is noted $premise(g)$. Notice that, according to the previous assumptions, we consider that the agent knows how to achieve its goals (in G_a)—the agent knows which are the pieces of information it needs to achieve its goals. This means that the goals for which the agent does not know the information necessary to achieve them are not considered.

Definition 1. *Let C and C' be two conjunctions of literals. C is included in C', noted $C \subseteq C'$, iff all the literals of C are literals of C'. C is equal to C', noted $C = C'$, iff the literals of C are exactly the same as the literals of C'. The result of the intersection between C and C', noted $C \cap C'$, contains literals which are both in C and in C'. The result of the difference between C and C', noted $C \setminus C'$, contains literals which are in C but not in C'. The cardinality of a conjunction of literals C, noted $|C|$, corresponds to the number of literals in C.*

Definition 2. *Let S and S' be two sets of conjunctions of positive literals. $S \preceq^1 S'$ iff (i) $|S| \leq |S'|$ and (ii) if $|S| = |S'|$ then there is a bijection $f : S \rightarrow S'$ such that: $\forall \psi \in S \ \psi \subseteq f(\psi)$. $S \prec^1 S'$ iff $S \preceq^1 S'$ and $S' \not\preceq^1 S$.*

Definition 3. *Let S and S' be two sets of conjunctions of positive literals. $S \preceq^2 S'$ iff (i) $|S| \leq |S'|$ and (ii) if $|S| = |S'|$ then there is a bijection $f : S \rightarrow S'$ such that: $\forall \psi \in S \ |\psi| \leq |f(\psi)|$. $S \prec^2 S'$ iff $S \preceq^2 S'$ and $S' \not\preceq^2 S$.*

Thus $S \preceq^1 S'$ (resp., $S \preceq^2 S'$) iff S does not have more elements than S'; if S and S' have the same number of elements, then the conjunctions in S are included in the conjunctions of S' (resp., are shorter than those of S'). Notice that $\preceq^1$ is a preorder but it is not total. Some sets of conjunctions are incomparable, such as $\{p, q \wedge r\} \not\preceq^1 \{r, s\}$ and $\{r, s\} \not\preceq^1 \{p, q \wedge r\}$. $\preceq^2$ is a total preorder.

Lemma 1. [3] *Let S and S' be two sets of conjunctions of positive literals.*

- *If $S \preceq^1 S'$ then $S \preceq^2 S'$*
- *If $S \preceq^1 S'$ and $S' \preceq^1 S$ then $S = S'$*
- *If $S \preceq^2 S'$ and $S' \preceq^2 S$ then $S = S'$*
- *$S \prec^1 S'$ iff (i) $|S| < |S'|$ or (ii) $|S| = |S'|$ and there is a bijection $f : S \to S'$ such that $\forall \psi \in S \ \psi \subseteq f(\psi)$ and $\exists \psi_0 \in S \ \psi_0 \subset f(\psi_0)$.*
- *$S \prec^2 S'$ iff (i) $|S| < |S'|$ or (ii) $|S| = |S'|$ and there is a bijection $f : S \to S'$ such that $\forall \psi \in S \ |\psi| \leq |f(\psi)|$ and $\exists \psi_0 \in S \ |\psi_0| < |f(\psi_0)|$.*

Definition 4 (Missing Information). *Let a be an agent with its belief base B_a and its goal set G_a. Let $g \in G_a$ be such that $B_a \not\models g$. $Missing(B_a, g)$, is defined as follows:*

$$Missing(B_a, g) = \bigwedge_{l:l \in premise(g) \ and \ B_a \not\models l} l$$

$Missing(B_a, g)$ is the conjunction of all the literals in the premise of g which cannot be deduced from B_a (i.e., which are not yet believed by the agent). Therefore, in the particular case in which $B_a^m = \emptyset$, $Missing(B_a, g) = premise(g)$, i.e., the missing piece of information to achieve g is $premise(g)$.

Notice that the notion of missing information is defined only for the goals that are not already achieved (i.e, goals such that $B_a \not\models g$). A missing information associated to a goal is then the conjunction of all the literals representing the information need to achieve that goal (not yet achieved), and only these ones. Moreover, we would like to stress out that, according to Definition 4, the formula whose conclusion is g can be written as: $Missing(B_a, g) \wedge \psi_{B_a, g} \to g$ with $\psi_{B_a, g} \in L \setminus L_G$, $B_a \models \psi_{B_a, g}$ and $B_a \not\models Missing(B_a, g)$.

Proposition 1.

- *Let $\varphi \in L \setminus L_G$ be a formula and $g \in G_a$ be a goal of agent a. We have that $Missing(B_a \cup \varphi, g) \subseteq Missing(B_a, g)$.*
- *If $\psi \models \varphi$ then $Missing(B_a \cup \psi, g) \subseteq Missing(B_a \cup \varphi, g)$.*
- *Let $\varphi_1 \in L \setminus L_G$, $\varphi_2 \in L \setminus L_G$ be two formulas and $g \in G_a$ be a goal of agent a. We have that $Missing(B_a \cup (\varphi_1 \wedge \varphi_2), g) = Missing((B_a \cup \varphi_1) \cup \varphi_2), g)$.*

Definition 5 (Multiset of missing information). *Let a be an agent whose belief base is B_a and whose goal set is G_a. The multiset[4] of missing information to achieve the goals in G_a is: $Missing(B_a, G_a) = \{Missing(B_a, g_1), \ldots, Missing(B_a, g_k)\}$ with $\{g_1, \ldots, g_k\} = \{g_i \in G_a \ and \ B_a \not\models g_i\}$.*

---

[3] Proofs are omitted due to length limitation.

[4] Reminder: a multiset is a set whose elements can have several occurrences, such as $\{p, q, p\}$.

There is therefore as much missing information as there are unachieved goals, i.e., the cardinality of $Missing(B_a, G_a)$ corresponds to the number of goals that are not yet achieved. Actually, we would like to take into account the weight of the missing information and not only the number of missing information.

Example 1. Let us consider a propositional language whose letters are: p, q, r, g_1 and g_2 respectively meaning "I know the main papers about modal logic", "I know the main papers about BDI agents", "I know the Python language", "I can start writing the state of the art" and "My Monday's class is prepared". Let us consider $G_a = \{g_1, g_2\}$ and $B_a = \{p\} \cup \{p \wedge q \rightarrow g_1, r \rightarrow g_2\}$. We have that, $Missing(B_a, g_1) = q$, $Missing(B_a, g_2) = r$ and therefore, $Missing(B_a, G_a) = \{q, r\}$. This means that, in order to achieve its goals, the agent lacks knowledge about BDI agents and about the Python language.

The following proposition shows that adding a belief to the belief base B_a does not increase the number of missing conjunctions. Moreover, if this does not reduce it either, then it does not increase their size. Finally, if adding a belief to the belief base B_a reduces the number of missing conjunctions, then this means that such new belief allows to achieve one or more goals.

Proposition 2. *For all formula (piece of information) $\varphi \in L \setminus L_G$, we have:*

- $|Missing(B_a \cup \varphi, G_a)| \leq |Missing(B_a, G_a)|$.
- $\forall \varphi$ *if* $|Missing(B_a \cup \varphi, G_a)| = |Missing(B_a, G_a)|$ *then there is a bijection* $f : Missing(B_a \cup \varphi, G_a) \rightarrow Missing(B_a, G_a)$ *such that* $\forall \psi \in Missing(B_a \cup \varphi, G_a)$ $\psi \subseteq f(\psi)$.
- *If* $|Missing(B_a \cup \varphi, G_a)| < |Missing(B_a, G_a)|$ *then* $\exists G_i \in G_a$ *such that* $Missing(B_a, g_i) \in Missing(B_a, G_a)$ *and* $B_a \cup \varphi \models g_i$.

3   A Binary Approach

In this section, we characterize useful information for an agent in view of achieving its goals in two different ways. According to this binary approach, a piece of information is useful or not.

Definition 6. *Let a be an agent with its belief base B_a and its set of goals G_a. Formula $\varphi \in L \setminus L_G$ is U^1-useful for agent a iff $Missing(B_a \cup \varphi, G_a) \prec^1 Missing(B_a, G_a)$. We use the notation $U^1_{G_a, B_a} \varphi$ or, more simply, $U^1 \varphi$, when there is no ambiguity.*

According to this definition, a formula φ in $L \setminus L_G$ is useful for a in view of achieving its goals G_a iff being aware of φ allows a to reduce its information need either by reducing the number of missing conjunctions or by simplifying them. Restricting useful information to formulas of $L \setminus L_G$ only amounts (i) to restrict to information the agent must acquire in order to achieve its goals and (ii) to rule out the fact that a goal can be achieved by a other than through the acquisition of information recommended in the formulas whose aims are the conclusions.

Definition 7. *Let a be an agent with its belief base B_a and its goals G_a. The formula $\varphi \in L \setminus L_G$ is U^2-useful for a iff $Missing(B_a \cup \varphi, G_a) \prec^2 Missing(B_a, G_a)$. We use the notation $U^2_{G_a, B_a} \varphi$ or $U^2 \varphi$ when there is no ambiguity.*

According to this second definition, a formula φ of $L \setminus L_G$ is U^2-useful for a if knowing φ allows a to reduce its information need either by reducing the number of missing conjunctions or by reducing their size. However, the two previous definitions, based on different pre-orders, are equivalent as shown by the following proposition.

Proposition 3.
$$U^1 \varphi \iff U^2 \varphi.$$

Since $U^1 \varphi$ and $U^2 \varphi$ are equivalent, we will use just $U\varphi$ to denote both.

Example 2. **Example 1 (continued)** $Missing(B_a, G_a) = \{q, r\}$. $Missing(B_a \cup \{r\}, G_a) = \{q\}$. $Missing(B_a \cup \{q\}, G_a) = \{r\}$. $Missing(B_a \cup \{q \wedge r\}, G_a) = \emptyset$. Therefore, Ur, Uq and $U(q \wedge r)$. In addition, if x is a propositional letter of the language, we have $U(r \wedge x)$ which means that $r \wedge x$ is useful. Indeed, *knowing Python and Java* is useful for the agent because it allows the agent to achieve G_2.

The last remark in this example shows a limitation of this binary model. Indeed, r is useful and so is $r \wedge x$ because, like r, it reduces the agent's need for information. However, this could be questionable because $r \wedge x$ contains x, which does not result in reducing the agent's need for information. In other words, reading a document on Python and Java, certainly allows the agent to acquire useful information about Python to prepare the class, but leads the agent to read content about Java, not useful for achieving its goals. This limitation is emphasized by the following proposition.

Proposition 4. *Let φ_1 and φ_2 be two formulas of $L \setminus L_G$. If $U\varphi_1$ then $U(\varphi_1 \wedge \varphi_2)$.*

Some more results are given below.

Proposition 5.

- *If φ is not useful then $Missing(B_a \cup \varphi, G_a) = Missing(B_a, G_a)$*
- *If $\exists \psi \in Missing(B_a, G_a)$ such that $\varphi \models \psi$ then $U\varphi$.*
- *$\exists \varphi \, U\varphi \not\Rightarrow Missing(B_a, G_a) \models \varphi$*
- *$\exists \varphi \, Missing(B_a, G_a) \models \varphi \not\Rightarrow U\varphi$*

The first point of this proposition shows that adding unnecessary information to the agent's belief base does not change missing information. The second point shows that any information that implies missing information is useful. In particular, any missing information is useful. The reverse is obviously not true. See example 2: $r \wedge x$ is useful but does not belong to $Missing(B_a, G_a)$. Therefore, all missing information is useful, but some useful information is not missing. The third point illustrates the fact that useful information is not necessarily a logical consequence of the $Missing(B_a, G_a)$ set. Finally, the fourth point illustrates the fact that there are logical consequences of $Missing(B_a, G_a)$ set that are not useful.

4 An Ordinal Approach

In this section we are interested in a notion of *relative usefulness* by defining, in two different ways, a pre-order between the formulas. To compare two formulas φ_1 and φ_2, we compare the two sets of information that is missing once the piece of information is added to the belief base, i.e., we compare $Missing(B_a \cup \varphi_1, G_a)$ and $Missing(B_a \cup \varphi_2, G_a)$, by using either of the pre-orders $\preceq^1$ and $\preceq^2$. Here, the obtained definitions will not be equivalent (see Example 3).

Definition 8. *Let a be an agent, B_a be its belief base and G_a be its set of goals. Let φ_1 and φ_2 be two formulas of $L \setminus L_G$. φ_1 is at least as useful for a as φ_2, denoted by $\varphi_2 \preceq_u^1 \varphi_1$, iff $Missing(B_a \cup \varphi_1, G_a) \preceq^1 Missing(B_a \cup \varphi_2, G_a)$. φ_1 is strictly more useful for a than φ_2, denoted by $\varphi_2 \prec_u^1 \varphi_1$, iff $\varphi_2 \preceq^1 \varphi_1$ and $\varphi_1 \not\preceq^1 \varphi_2$. Finally, φ_1 is as useful for a as φ_2, denoted by $\sim_u^1$, iff $\varphi_2 \preceq_u^1 \varphi_1$ and $\varphi_1 \preceq_u^1 \varphi_2$.*

According to this definition, if one piece of information allows to achieve more goals than another, then it is more useful. If it makes it possible to achieve the same number of goals but if, for at least one goal, it makes it possible to reduce missing information, then it is more useful.

Obviously, $\varphi_2 \prec_u^1 \varphi_1$ iff $Missing(B_a \cup \varphi_1, G_a) \prec^1 Missing(B_a \cup \varphi_2, G_a)$ and $\varphi_2 \sim_u^1 \varphi_1$ iff $Missing(B_a \cup \varphi_1, G_a) = Missing(B_a \cup \varphi_2, G_a)$. $\preceq_u^1$ is a pre-order on all the propositional formulas but not a total pre-order. For example, in Example 3 below, $p \wedge q$ and $p \wedge r$ are incomparable. Indeed $Missing(B_a \cup (p \wedge q), G_a) = \{r\}$ and $Missing(B_a \cup (p \wedge r), G_a) = \{q\}$ and $\{r\} \not\preceq^1 \{q\}$ and $\{q\} \not\preceq^1 \{r\}$.

Definition 9. *Let a be an agent, B_a be its belief base and G_a be its set of goals. Let φ_1 and φ_2 be two formulas of $L \setminus L_G$. φ_1 is at least as useful for a as φ_2, denoted by $\varphi_2 \preceq_u^2 \varphi_1$, iff $Missing(B_a \cup \varphi_1, G_a) \preceq^2 Missing(B_a \cup \varphi_2, G_a)$. φ_1 is strictly more useful for a than φ_2, denoted by $\varphi_2 \prec_u^2 \varphi_1$, iff $\varphi_2 \preceq_U^2 \varphi_1$ and $\varphi_1 \not\preceq_U^2 \varphi_2$. Finally, φ_1 is as useful for a as φ_2, denoted by $\sim_u^2$, iff $\varphi_2 \preceq_u^2 \varphi_1$ and $\varphi_1 \preceq_u^2 \varphi_2$.*

According to this definition, if one piece of information allows to achieve more goals than another, then it is more useful. If it achieves the same number of goals and if the missing information is generally shorter, then it is more useful.These two definitions are not equivalent as shown below.

Example 3. Let us suppose that: $B_a = \{p \wedge q \to g_1, p \wedge r \to g_2\}$ and $G_a = \{g_1, g_2\}$. We have for instance, $Missing(B_a \cup (p \wedge x), G_a) = \{q, r\}$ and $Missing(B_a \cup r, G_a) = \{p \wedge q, p\}$. Thus $r \prec_u^2 (p \wedge x)$ but $r \not\prec_u^1 (p \wedge x)$.

Proposition 6. *If $\psi \models \varphi$ then $\varphi \preceq_U^1 \psi$ and $\varphi \preceq_U^2 \psi$.*

In particular $\varphi_1 \preceq_U^1 \varphi_1 \wedge \varphi_2$ and $\varphi_1 \preceq_U^2 \varphi_1 \wedge \varphi_2$. That is to say $\varphi_1 \wedge \varphi_2$ is at least as useful, in the sense of $\preceq_U^1$ (and of $\preceq_U^2$) than φ_1. However, we do not have $\varphi_1 \prec_U^1 \varphi_1 \wedge \varphi_2$ neither $\varphi_1 \prec_U^2 \varphi_1 \wedge \varphi_2$ as shown in the previous examples where $p \sim^1 p \wedge x$ and $p \sim^2 p \wedge x$.

5   A Numerical Approach

In this section, we follow a numerical approach by associating each piece of information with a usefulness degree. To begin with, we state some rationality postulates such a measure must satisfy. The general case will not be treated, and we will limit ourselves to calculating the degree of usefulness *of conjunctions of positive literals*. Let φ be a conjunction of positive literals. We define:

- $Cons(B_a, \varphi) = \{l\ positive\ literal\ of\ L \setminus L_G : B_a \cup \varphi \models l\}$
- $N_1(\varphi) = \Sigma_{g \in G_a} |Cons(B_a, \varphi) \cap Missing(B_a, g)|$
- $N_2(\varphi) = \Sigma_{g \in G_a} |Missing(B_a, g) \setminus Cons(B_a, \varphi)|$
- $N_3(\varphi) = |\varphi \setminus \cup_{G \in G_a} Missing(B_a, g)|$

$Cons(B_a, \varphi)$ is the set of all the positive literals that are deducible after adding φ to B_a. $N_1(\varphi)$ counts the positive literals common to $Cons(B_a, \varphi)$ and to the missing information. The larger the $N_1(\varphi)$, the more φ reduces the missing information to achieve the goals. $N_2(\varphi)$ counts the positive literals of missing information that are not in $Cons(B_a, \varphi)$.
Notice that $N_2(\varphi) = \Sigma_{g \in G_a} |Missing(B_a, g)| - N_1(\varphi)$. Therefore, if $N_1(\varphi)$ increases, $N_2(\varphi)$ decreases. $N_3(\varphi)$ counts the positive literals of φ that are not literals of missing information. Adding them is therefore not useful to achieve the goals.

Let us consider again agent a whose belief base is B_a and goal set is G_a.

Definition 10. *The set of goals that a formula φ allows the agent to achieve is:*

$$E_{B_a, G_a}(\varphi) = \{g \in G_a, B_a \not\models g\ and\ B_a \cup \varphi \models g\}$$

We use $E(\varphi)$ when there is no ambiguity.

Let $U(\varphi)$ be a real number representing how much φ is useful for a. We have based our definition on the following postulates.

Monotonicity on the number of goals:

(P1) $|E(\varphi_1)| < |E(\varphi_2)| \implies U(\varphi_1) < U(\varphi_2)$.

The number of goals that a piece of information allows an agent to achieve should influence the degree of usefulness of such a piece of information for the agent. Intuitively, a piece of information which allows to achieve a higher number of goals (with respect to another piece of information) should be more useful.

Monotonicity on the quantity of information needed:

(P2) $|E(\varphi_1)| = |E(\varphi_2)|$ and $N_1(\varphi_1) > N_1(\varphi_2) \implies U(\varphi_1) > U(\varphi_2)$.

The amount of missing information (needed information) provided by a formula should influence its degree of usefulness. Intuitively, when two formulas allow to achieve the same number of goals, one of them is more useful than the other if it reduces the amount of missing information more than the other.

Monotonicity on the quantity of useless information:

(P3) $|E(\varphi_1)| = |E(\varphi_2)|$ and $N_1(\varphi_1) = N_1(\varphi_2)$ and $N_3(\varphi_1) < N_3(\varphi_2) \Longrightarrow U(\varphi_1) > U(\varphi_2)$.

The amount of useless information conveyed by a piece of information should also influence its degree of usefulness. Intuitively, when two formulas allow to achieve the same number of goals, a formula is more useful than another if it provides less useless information than the other. This idea agrees with the maxmin principle of Grice's. Useless information while not being harmful in view of reaching a goal may produce an overhead on whom has to process it which may be qualified as a cost.

Equality:

P4 $|E(\varphi_1)| = |E(\varphi_2)|$ and $N_1(\varphi_1) = N_1(\varphi_2)$ and $N_3(\varphi_1) = N_3(\varphi_2) \Longrightarrow U(\varphi_1) = U(\varphi_2)$.

Two pieces of information which allow to achieve the same number of goals, and which have exactly the same amount of useful and useless information should have the same degree of usefulness.

To sum up, according to **(P1)**, the higher the number of goals that a formula makes it possible to achieve, the higher its usefulness degree. According to **(P2)**, **(P3)** and **(P4)**, when two formulas allow to achieve the same number of goals (whether the goals are the same, different or even no goals at all), then the more a formula reduces missing information the more useful it is. Moreover, in case of equality, the most useful information is the one which brings the least useless information; finally, if they have the same number of useless pieces of information, then they have the same usefulness degree. These postulates are consistent because their premises are incompatible.

Notice that, according to these postulates, if $N_1(\varphi_1) = N_1(\varphi_2)$ and $N_3(\varphi_1) = N_3(\varphi_2)$ then $U(\varphi_1) = U(\varphi_2)$.

In the following, we provide the definition of a usefulness measure U which satisfies these postulates.

Definition 11. *Let a be an agent whose goals are in G_a and let φ be a conjunction of positive literals. We define the usefulness degree[5] by:*

$$U(\varphi) = \frac{1}{|G_a|+1} \left[|E(\varphi)| + \frac{N_1(\varphi)}{N_1(\varphi)+N_2(\varphi)+\frac{N_3(\varphi)}{N_3(\varphi)+1}} \right]$$

$N_1(\varphi)$ quantifies the useful part of φ for the agent, while $N_2(\varphi)$ quantifies the agent's disappointment (lack of needed information) towards φ and, finally, $N_3(\varphi)$ quantifies the disturbance caused to the agent by the unexpected and unnecessary content of φ. Our definition of usefulness takes these three aspects into consideration.

---

[5] Such a degree should be noted $U_{B_a,G_a}(\varphi)$ but we will note it $U(\varphi)$ when there is no ambiguity.

The intuitive idea behind this definition of usefulness is as follows. The usefulness of information can be seen as a calculation of the similarity between the information the agent needs to achieve its goals and the piece of information that arrives. The more direct or indirect elements (that can be deduced) there are in common between the two, the more useful the information will be. We would like to stress that this fact allows to account for the serendipity factor [17] in the definition of usefulness. Indeed, an agent gets (asks for) a piece of information to achieve a given goal, but if the received piece of information also helps achieve other goals then this fact is considered in the computation of the usefulness. However, the number of common elements is not always enough to distinguish the degrees of usefulness between two pieces of information. Indeed, in some cases it would also be necessary to take into account their differences. We have been inspired by Tversky's idea [18], according to which, in order to calculate the similarity between two objects A and B, we should consider, in addition to what they have in common, what distinguishes them, i.e., the features of A which are not features of B and vice-versa. This is the reason why we have considered these three values, $N_1(\varphi)$, $N_2(\varphi)$ and $N_3(\varphi)$, in our definition.

Remark 1 *We can notice that for any conjunction of positive literals φ we have:*

$$0 \leq \frac{N_1(\varphi)}{N_1(\varphi)+N_2(\varphi)+\frac{N_3(\varphi)}{N_3(\varphi)+1}} \leq 1.$$

$$\frac{N_1(\varphi)}{N_1(\varphi)+N_2(\varphi)+\frac{N_3(\varphi)}{N_3(\varphi)+1}} = 1 \implies E(\varphi) = G_a.$$

Proposition 7. *The measure $U(\varphi)$ proposed in Definition 11 satisfies postulates* **(P1)– (P4)**.

Example 4. Let us consider: $B_a = \{q\} \cup \{p \wedge q \to g_1, p \wedge r \to g_2\}$ and $G_a = \{g_1, g_2\}$. We have then $Missing(B_a, G_a) = \{p, p \wedge r\}$. We obtain $U(p \wedge r) = 1$, $U(p) = 5/6$, $U(p \wedge q) = U(p \wedge x) = 11/14$, $U(r) = 1/6$, $U(q \wedge r) = 1/7$, $U(q \wedge x) = U(q) = 0$. In other words, $p \wedge r$ is the piece of information that has the maximal degree of usefulness, which is explained by the fact that adding $p \wedge r$ allows to achieve both goals g_1 and g_2. The usefulness of p is lower than the usefulness of $p \wedge r$ but it is higher than those of the other formulas, because adding p allows to achieve a goal (g_1). On the other hand, $p \wedge q$, is less useful than p because of q: the agent already knows q therefore q is not useful anymore for the agent because not novel. The same reasoning holds for $p \wedge x$. Formula r instead is less useful because it only reduces missing information regarding one single goal. It is easy to understand that $q \wedge r$ is less useful than r once more because of the unnecessary information q. Obviously, q and $q \wedge x$ are not useful at all because they do not help progressing towards a goal.

Proposition 8. *If $U(\varphi_1) = U(\varphi_2)$ then $|E(\varphi_1)| = |E(\varphi_2)|$ and $N_1(\varphi_1) = N_1(\varphi_1)$, $N_2(\varphi_1) = N_2(\varphi_1)$, $N_3(\varphi_1) = N_3(\varphi_1)$.*

By this proposition, the only way two formulas can have the same usefulness is by having the same values for these three parameters. This shows that definition 11 does not permit any compensation: a variation of one of these three values cannot be compensated by the variation of the others.

Particular Cases

- **When $B_a^m = \emptyset$:** In the case where $B_a^m = \emptyset$, i.e., when the only beliefs of the agent concern the agent's needs in terms of information about the way to achieve its goals, we have: $Missing(B_a, g) = premise(g)$ and $Cons(B_a, \varphi) = \varphi$. $U(\varphi)$ can then be written as:

$$U(\varphi) = \frac{1}{|G_a+1|} \cdot \left[|E(\varphi)| + \frac{N_1(\varphi)}{K + \frac{N_3(\varphi)}{N_3(\varphi)+1}} \right]$$

  with $E(\varphi) = \{g \in G_a, premise(g) \subseteq \varphi\}$, $N_1(\varphi) = \Sigma_{G \in G_a} |\varphi \cap Premise(g)|$, $K = \Sigma_{g \in G_a} |premise(g)|$, $N_3(\varphi) = |\varphi \setminus \cup_{g \in G_a} Premise(g)|$

- **When $B_a^m = \emptyset$ and G_a is a singleton:** In this case, the agent has a single goal, g_0, and its only beliefs is the formula which expresses the information need for achieving that single goal. $U(\varphi)$ can then be written as follows:

$$U(\varphi) = \frac{1}{2} \cdot \left(n(\varphi) + \frac{|\varphi \cap premise(g_0)|}{|premise(g_0)| + \frac{|\varphi \setminus premise(g_0)|}{|\varphi \setminus premise(g_0)|+1}} \right)$$

  with $n(\varphi) = 1$ if $premise(g_0) \subseteq \varphi$ and $n(\phi) = 0$ otherwise.

Example 5. Take $premise(g_0) = a \wedge b$. Then we have $U(c) = 0$, $U(a \wedge c) = 1/5$, $U(a) = 1/4$, $U(a \wedge b \wedge c) = 9/10$, $U(a \wedge b) = 1$. In other words, c is not useful at all because knowing c does not allow the agent to reach or get closer to its goal. $a \wedge c$ is a little more useful, because even if knowing c is not useful to the agent, knowing a allows it to get a little closer to its goal. a is more useful than $a \wedge c$ because it does not add unnecessary information. $a \wedge b \wedge c$ is even more useful because even if it adds unnecessary information, it allows the agent to achieve its goal. Finally, $a \wedge b$ is the most useful because it allows the agent to reach its goal and does not add any unnecessary information.

6   An Example of Application to Information Retrieval

In this section, we will first recall some relevance dimensions in information retrieval which have been used in the literature [8] to propose documents to a user (who now takes the place of what we called "agent" in the above general framework). We will then compare those dimensions with the usefulness measure we are proposing here. However, to have a fair comparison, we need to reformulate those dimensions in a logical setting [1].

6.1   A Refresher on Relevance Dimensions

The *aboutness* [5] dimension is a core notion in Information Retrieval. It is used to compute the topical matching between a document and a user query. However, its modeling gave raise to several distinct interpretations, which characterize a variety of Information Retrieval models, of which the *vector space model* is an example. Formally, in the vector space model, a piece of information or, more generally, a document d, can be

represented as a vector of T elements, $d = [w_{1d}, \ldots, w_{|T|d}]$. The user interests are represented by a vector $q = [w_{1q}, \ldots, w_{|T|q}]$, $|T|$ being the size of the term vocabulary used. Different choices have been made in the literature regarding the values of w_{id}, for example: simply based on the presence or absence of a word in the document, in this case the vector contains values in $\{0, 1\}$, or based on the frequency of the word in the document and in the whole repository (TF-IDF) [2]. Here, we will use the vector space model interpretation, and, like in [8], in addition to the aboutness measure, we will consider the *appropriateness* dimension (proposed in [8]) and the *coverage* dimension (proposed in [13]). We have considered those three relevance dimensions because they explicitly account for the user query/goals. This is not the case for the *popularity* relevance dimension for example.

Aboutness The term *a*boutness (*t*opical relevance) is formally defined as follows. Let $d = [w_{1d} \ldots w_{|T|d}]$ and $q = [w_{1q} \ldots w_{|T|q}]$ representing document d and query q, respectively, with $|T|$ representing the size of the term vocabulary used. The measure of *aboutness* (*t*opical relevance) is calculated by the standard cosine-similarity [15]:

$$\text{Aboutness}_{IR}(d, q) = \frac{\sum_{i=1}^{|T|} (w_{iq}.w_{id})}{\sqrt{\sum_{i=1}^{|T|} w_{iq}^2 \cdot \sum_{i=1}^{|T|} w_{id}^2}}. \tag{1}$$

Coverage The *coverage* criterion is assessed on the document representation and on the user profile representation. It measures how strongly the user interests are included in a document.

$$\text{Coverage}_{IR}(d, q) = \frac{\sum_{i=1}^{|T|} \min(w_{iq}, w_{id})}{\sum_{i=1}^{|T|} w_{iq}}. \tag{2}$$

This function produces the maximum value 1 when the non null elements in q's vector also belong to d's vector. It produces the value zero when the two vectors have no common element. Moreover, the value of the function increases with the increase of the number of common elements.

Appropriateness This dimension allows to measure how appropriate or how seemly a document is with respect to the user interests.

$$\text{Appropriateness}_{IR}(d, q) = 1 - \frac{\sum_{i=1}^{|T|} |w_{iq} - w_{id}|}{|T|}. \tag{3}$$

According to this definition, a piece of information is considered *fully appropriate* if it covers all the user interests. However, if in addition it covers other subjects, it is considered *less appropriate*.

6.2   Reformulation in Logic

We can consider a user query in information retrieval as the information needed to achieve a goal. This way, the premise of the goal can be represented by a formula that

corresponds to the agent's information needed. Let φ and ψ be two conjunctions of positive literals of a propositional language. We have:

$$\text{Aboutness}_{Logic}(\varphi, \psi)) = \frac{|\varphi \cap \psi|}{\sqrt{|\varphi| \cdot |\psi|}},$$

$$\text{Coverage}_{Logic}(\varphi, \psi) = \frac{|\varphi \cap \psi|}{|\psi|},$$

$$\text{Appropriateness}_{Logic}(\varphi, \psi) = 1 - \frac{|\varphi \backslash \psi| + |\psi \backslash \varphi|}{|L|}.$$

After replacing the premises of the agent's goal by the formula ψ, the measure defined in definition 11 is then re-written as follows:

$$U(\varphi, \psi) = \frac{1}{2} \cdot \left(n(\varphi) + \frac{|\varphi \cap \psi|}{|\psi| + \frac{|\varphi \backslash \psi|}{|\varphi \backslash \psi| + 1}} \right)$$

with $n(\varphi) = 1$ if $\psi \subseteq \varphi$ and $n(\phi) = 0$ otherwise.

More precisely, we consider a propositional language L that has $|T|$ propositional letters $p_1 \ldots p_{|T|}$ and a letter g_0 representing the goal of the user. A document d can then be represented by a formula noted φ_d defined as: $\varphi_d = \bigwedge_{i=1,\ldots,|T| \ and \ w_{i,d}=1} p_i$. A query q can also be represented by a formula noted $premise(g_0)$ defined by $\psi_q = \bigwedge_{i=1,\ldots,|T| \ and \ w_{i,q}=1} p_i$.

The following proposition allows us to reformulate in logic the three IR relevance dimensions we have considered from the literature.

Proposition 9.

$$\text{Aboutness}_{IR}(d, q) = \text{Aboutness}_{Logic}(\varphi_d, \psi_q)$$
$$\text{Coverage}_{IR}(d, q) = \text{Coverage}_{Logic}(\varphi_d, \psi_q)$$
$$\text{Appropriateness}_{IR}(d, q) = \text{Appropriateness}_{Logic}(\varphi_d, \psi_q)$$

Example 6. Let us consider again Example 5, with the propositional language whose letters are a, b, c and g_0. $\psi = a \wedge b$ and let us consider the five formulas: $\varphi_1 = c$, $\varphi_2 = a \wedge c$, $\varphi_3 = a$, $\varphi_4 = a \wedge b \wedge c$, and $\varphi_5 = a \wedge b$. The following table summarizes the values of the four measurements.

φ	$About$	Cov	$Approp$	U
$\varphi_1 = c$	0	0	0	0
$\varphi_2 = a \wedge c$	1/2	1/2	1/3	1/5
$\varphi_3 = a$	$\frac{1}{\sqrt{2}}$	1/2	2/3	1/4
$\varphi_4 = a \wedge b \wedge c$	$\frac{2}{\sqrt{6}}$	1	2/3	9/10
$\varphi_5 = a \wedge b$	1	1	1	1

A number of observations emerge from these results. First of all, we notice that two formulas can have identical degrees of coverage without their degrees of usefulness being identical. Thus, $Coverage(\varphi_2, \psi) = Coverage(\varphi_3, \psi)$ but $U(\varphi_2, \psi) \neq$

$U(\varphi_3, \psi)$. Similarly, two formulas may have identical degrees of appropriateness without their degrees of usefulness being identical. Thus, $Appropriatemess(\varphi_3, \psi) = Appropriateness(\varphi_4, \psi)$ but $U(\varphi_3, \psi) \neq U(\varphi_4, \psi)$. We also notice that a and $a \wedge b \wedge c$ have identical appropriateness values although for different reasons: $appriopriateness(a, a \wedge b) = 2/3$ because a says nothing about b, whereas this is part of the user's information need, and $appriopriateness(a \wedge b \wedge c, a \wedge b) = 2/3$ because $a \wedge b \wedge c$, although providing all the information the user need to achieve his/her goal, it provides unnecessary information, c. On the other hand, these different reasons lead to different degrees of usefulness and, in particular, $U(a, a \wedge b)$ is much lower than $U(a \wedge b \wedge c, a \wedge b)$. Indeed, by definition, U favors information that allows the user need to be satisfied (this is fully the case with $a \wedge b \wedge c$ whereas it is partially the case with a). Even if $a \wedge b \wedge c$ provides unnecessary information, namely c, the user will be able to achieve his/her goal with it, unlike with a.

The following proposition provides some comparisons between the U measure and the IR ones.

Proposition 10. *Let φ and ψ be two conjunctions of literals.*

- $U(\varphi, \psi) = \mathrm{Aboutness}(\varphi, \psi) = \mathrm{Appropriateness}(\varphi, \psi) = 1 \Longleftrightarrow \varphi = \psi$.
- $\mathrm{Coverage}(\varphi, \psi) = 1 \Longleftrightarrow \psi \subseteq \varphi$.
- $U(\varphi, \psi) = \mathrm{Aboutness}(\varphi, \psi) = \mathrm{Appropriateness}(\varphi, \psi) = \mathrm{Coverage}(\varphi, \psi) = 0 \Longleftrightarrow \varphi \cap \psi = \emptyset$.
- $\mathrm{Coverage}(\varphi_1, \psi) < \mathrm{Coverage}(\varphi_2, \psi) \Longrightarrow U(\varphi_1, \psi) < U(\varphi_2, \psi)$
- $\mathrm{Appropriateness}(\varphi_1, \psi) \leq \mathrm{Appropriateness}(\varphi_2, \psi)$ *and* $\mathrm{Coverage}(\varphi_1, \psi) = \mathrm{Coverage}(\varphi_2, \psi) \Longrightarrow U(\varphi_1, \psi) \leq U(\varphi_2, \psi)$.

7   Conclusion and Future Work

We have proposed three approches to define the notion of usefulness for a cognitive agent. A *binary approach*, which allows to classify a piece of information as being *useful or not*. An *ordinal approach*, which allows to compare two pieces of information in order to establish which one is more useful. Two different operators have been proposed in this case: a pre-order operator and a total order operator. However, and like for the binary approach, the proposed ordinal approach does not allow to consider unnecessary information. This is accounted for by the third approach by means of a numerical definition of usefulness. We have compared, through an easy to understand example, three IR measures from the literature with our numerical measure. The results of the comparison show that our numerical definition of usefulness, based on the cognitive aspects of the user, allows to capture in a single value different dimensions, without the need for eliciting an explicit priority order on the dimensions from the user. In addition, it allows to somehow account for the *serendipity* factor (see Example 4). Moreover, it also allows to account for *novelty* with respect to the user's beliefs, not only with respect to the past user interactions as usual in the literature (see again Example 4, in which the fact that a piece of information contains information already known by the user diminishes its usefulness).

An application of our framework that would be interesting to investigate is its use to reduce the needs to coordinate multiple *assistive agents* advising the same user [16]. Other possible applications would be in the case of the Information Flow Problem in multi-agent systems, in which there is a need to ensure an adequate exchange of information within a system [3], and in the case of BDI personal medical assistant agents, where one critical requirement is to (automatically) produce an accurate documentation [6].

We also plan to extend our framework in a more general case where premise of a goal is not restricted to a conjunction but may be a more complex formula. We also plan to consider weighted goals in order to take into account the importance of goals in the definition of information usefulness.

References

1. Karam Abdulahhad, Catherine Berrut, Jean-Pierre Chevallet, and Gabriella Pasi. Modeling information retrieval by formal logic: A survey. *ACM Comput. Surv.*, 52(1):15:1–15:37, 2019.
2. Ricardo A. Baeza-Yates and Berthier A. Ribeiro-Neto. *Modern Information Retrieval - the concepts and technology behind search, Second edition*. Pearson Education Ltd., Harlow, England, 2011.
3. Luis Búrdalo, Andrés Terrasa, Vicente Julián, and Ana García Fornes. The information flow problem in multi-agent systems. *Engineering Applications of Artificial Intelligence*, 70:130 – 141, 2018.
4. Charles L. A. Clarke, Maheedhar Kolla, Gordon V. Cormack, Olga Vechtomova, Azin Ashkan, Stefan Büttcher, and Ian MacKinnon. Novelty and diversity in information retrieval evaluation. In *SIGIR*, pages 659–666. ACM, 2008.
5. William S. Cooper. A definition of relevance for information retrieval. *Information Storage and Retrieval*, 7(1):19–37, 1971.
6. Angelo Croatti, Sara Montagna, Alessandro Ricci, Emiliano Gamberini, Vittorio Albarello, and Vanni Agnoletti. BDI personal medical assistant agents: The case of trauma tracking and alerting. *Artificial Intelligence in Medicine*, 2018.
7. Michael da Costa Móra, José Gabriel Pereira Lopes, Rosa M. Vicari, and Helder Coelho. BDI models and systems: Bridging the gap. In *Proceedings of ATAL'98*, pages 11–27, 1998.
8. Célia da Costa Pereira, Mauro Dragoni, and Gabriella Pasi. Multidimensional relevance: A new aggregation criterion. In M. Boughanem, C. Berrut, J. Mothe, and C. Soulé-Dupuy, editors, *ECIR*, volume 5478 of *Lecture Notes in Computer Science*, pages 264–275. Springer, 2009.
9. Frédéric Flouvat, Jérémy Sanhes, Claude Pasquier, Nazha Selmaoui-Folcher, and Jean-François Boulicaut. Improving pattern discovery relevancy by deriving constraints from expert models. In *ECAI*, volume 263 of *Frontiers in Artificial Intelligence and Applications*, pages 327–332. IOS Press, 2014.
10. Herbert P. Grice. Logic and conversation. In Peter Cole and Jerry L. Morgan, editors, *Syntax and Semantics: Vol. 3: Speech Acts*, pages 41–58. Academic Press, New York, 1975.
11. Xiaoli Huang and Dagobert Soergel. Relevance: An improved framework for explicating the notion. *JASIST*, 64(1):18–35, 2013.
12. Jack Minker. An overview of cooperative answering in databases. In *FQAS*, volume 1495 of *Lecture Notes in Computer Science*, pages 282–285. Springer, 1998.

13. Gabriella Pasi, Gloria Bordogna, and Robert Villa. A multi-criteria content-based filtering system. In *SIGIR '07: Proceedings of the 30th annual international ACM SIGIR conference on Research and development in information retrieval*, pages 775–776, 2007.
14. Anand S. Rao and Michael P. Georgeff. Modeling rational agents within a BDI-architecture. In *KR'91*, pages 473–484, 1991.
15. Gerard Salton and Michael McGill. *Introduction to Modern Information Retrieval*. McGraw-Hill Book Company, 1984.
16. Budhitama Subagdja, Ah-Hwee Tan, and Yilin Kang. A coordination framework for multi-agent persuasion and adviser systems. *Expert Syst. Appl.*, 116:31–51, 2019.
17. Elaine G. Toms. Serendipitous information retrieval. In *DELOS*, 2000.
18. Amos Tversky. Features of similarity. *Psychological Review*, 84(4):327–352, 1977.
19. Ziqi Zhang, Johann Petrak, and Diana Maynard. Adapted textrank for term extraction: A generic method of improving automatic term extraction algorithms. *Procedia Computer Science*, 137:102 – 108, 2018. Proceedings of the 14th International Conference on Semantic Systems 10th–13th of September 2018 Vienna, Austria.