



A Framework for Comparative Analysis of Intention Mining Approaches

Rébecca Déneckère, Elena Kornyshova, Charlotte Hug

► To cite this version:

Rébecca Déneckère, Elena Kornyshova, Charlotte Hug. A Framework for Comparative Analysis of Intention Mining Approaches. International Conference on Research Challenges in Information Science (RCIS), May 2021, Limassol, Cyprus. pp.20-37, 10.1007/978-3-030-75018-3_2 . hal-03250113

HAL Id: hal-03250113

<https://hal.science/hal-03250113>

Submitted on 27 Sep 2021

HAL is a multi-disciplinary open access archive for the deposit and dissemination of scientific research documents, whether they are published or not. The documents may come from teaching and research institutions in France or abroad, or from public or private research centers.

L'archive ouverte pluridisciplinaire **HAL**, est destinée au dépôt et à la diffusion de documents scientifiques de niveau recherche, publiés ou non, émanant des établissements d'enseignement et de recherche français ou étrangers, des laboratoires publics ou privés.

A Framework for Comparative Analysis of Intention Mining Approaches

Rébecca Déneckère¹, Elena Kornyshova², and Charlotte Hug¹

¹ CRI, Université Paris 1 Panthéon Sorbonne, Paris, France

² CEDRIC, Conservatoire National des Arts et Métiers, Paris, France

rebecca.deneckere@univ-paris1.fr, elena.kornyshova@cnam.fr

Abstract. Intention Mining has the purpose to manipulate of large volumes of data, integrate information from different sources and formats and extract useful insights as facts from this data in order to discover users' intentions. It is used in different fields: Robotics, Network forensics, Security, Bioinformatics, Learning, Map Visualization, Game, etc. There is actually a large variety of intention mining techniques applied to different domains as information retrieval, security, robotics, etc. However, no systematic review had been conducted on this recent research domain. There is a need to understand what is Intention Mining, what is its purpose, what are the existing techniques and tools to mine intentions. In this paper, we propose a comparison framework to structure and to describe the domain of Intention Mining for a further complete systematic literature review of this field. We validate our comparison framework by applying it to five relevant approaches in the domain.

Keywords: Intention, Intention Mining, Comparison Framework.

1 Introduction

The concept of intention is becoming increasingly important in different areas of computer science as information retrieval, network forensics, security, robotics and bioinformatics, among others, to understand the goals of the users of the systems. [1] defines an intention as “a determination to act in a certain way; a concept considered as the product of attention directed to an object or knowledge”. Psychology specifies “Our common sense psychological scheme admits of intentions as states of mind; and it also allows us to characterize actions as done intentionally, or with a certain intention” [2]. These intentions can be clearly and explicitly stated or they can be implicitly expressed in natural language in different kind of sources as documents, queries, logs, etc. Many approaches tackle the problem of identifying intentions by using mining techniques.

This research domain is quite new, and the term *Intention Mining* (IM) has different meanings according to the communities. The proposed IM techniques and their aims are quite different from one domain to another. Moreover, *Intention Mining* is not the only term used to designate this activity as it has many synonyms: an “intention” may be an “intent”, a “goal”, an “objective”, etc.; mining may signify “discovery”, “analysis”, etc. It is then necessary to conduct a literature review on what is IM and how it is defined by

the different research communities in Computer Science context. There are techniques aiming at discovering intentions behind web queries to improve the recommended web pages or services [3-4]; others propose techniques to identify intentions in home videos to provide adapted home video services [5], while others propose to discover intentions hidden behind user activities defined in logs to understand users ways of working [6], etc. There is an obvious lack of overview of the existing IM techniques as they are defined in different areas.

Our future goal is to conduct a systematic literature review (SLR) on IM. To our knowledge, no proper systematic literature review has been conducted on Intention Mining. [7] propose a state of the art of intention recognition but it is not a systematic literature review. [8] proposes a review of intention process mining but restrict itself to the discovery of goal-oriented processes. Moreover, these works do not use a structured framework to analyse and compare the existing approaches.

To provide a detailed and structured overview of the IM approaches and to clarify IM elements and categories, we have elaborated a comparison framework composed of four dimensions: object, usage, method and tool. This structure was inspired from [9-12]. In this paper, we detail the comparison framework containing a set of criteria to differentiate and compare these different approaches and apply it to the most referenced papers of our study in order to check its feasibility. A detailed comparison framework is useful (i) for our future research, to carry out the SLR to compare the existing literature on IM and to define open issues, (ii) for a user, to be able to quickly compare the existing IM approaches and select the one best fitting its case, and (iii) for a new IM approach, to position it with regards to the existing literature.

The paper is organized as follows. Section 2 describes our research process. Section 3 presents the comparison framework to analyse the IM approaches. In Section 4, we apply the proposed framework to the 5 most referenced IM approaches and we conclude and define our research perspectives in Section 5.

2 Research Process

Our goal in this paper is to establish a comparison framework that will serve as a scope for the future systematic literature review. Our research process is presented on Fig. 1.

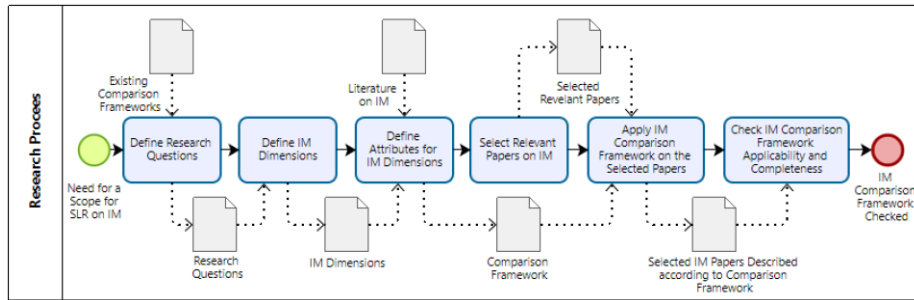


Fig. 1. Research Process Used to Develop the Intention Mining Comparison Framework.

Define Research Questions. The research goal is to identify a set of characteristics allowing to compare works on intention mining. The existing literature on existing comparison frameworks (applied to the field of software engineering [9], process engineering [10], decisional methods [11], method engineering [12], etc.) was analyzed in order to identify the research questions covering the main dimensions usually used to compare different approaches as approaches objects (*What is IM?*), approaches goals (*Why IM?*), approaches methods (*How is done IM?*), and, finally, approaches associated tools (*By which means is done IM?*). Thus, the defined research questions are:

- RQ1: What is intention mining? How IM is defined in the research papers or community, what does it consist in?
- RQ2: Why using intention mining? What is the purpose of IM in the research papers or community?
- RQ3: How is intention mining achieved? What are the existing IM techniques?
- RQ4: By which means is the intention mining method put in practice? What are the tools and algorithms that support the mining of intentions?

Define IM Dimensions. Based on these research questions, we have identified four IM dimensions of our comparison framework: Object dimension, Usage dimension, Method Dimension, and Tool Dimension (all detailed in the following section).

Define Attributes for IM Dimensions. The next step was to read the literature on IM. For this purpose, we used papers on IM appearing in Google Scholar. At this step, it was firstly important to identify various characteristics allowing to differentiate IM approaches. For instance, several approaches mention only intentions extraction [13], whereas other ones establish intentions models [6,14]. Thus, we identified the characteristic (an attribute in our comparison framework) *Structuredness*. Once different attributes identified, we grouped them into the four previously established dimensions according to their nature to structure the Comparison Framework.

Select Relevant Papers on IM. To select relevant papers, we have used the ongoing work related to SRL. Our main research query used the following combination of key words: (user OR actor) AND (goal OR intention OR intent OR objective) AND (discovery OR mining OR analysis). We have identified 238 research papers from the research bases IEEE, Springer, ACM, Science Direct with the following inclusion criteria: papers should be in French or in English, title or abstract should be related to IM. Then we applied some exclusion criteria (papers should be journal or conference papers, be present in DBLP (except for 2020 papers), and restrictions related to the citation number: if dated before 2010, we excluded papers cited less than 100 times, if dated between 2010 and 2017, we removed papers cited less than 10 times). We have finally obtained 145 papers. We have extracted the citation number from Google Scholar (access on January 2021) for each of these papers and selected the three most referenced papers [15], [16], and [3] having correspondingly 364, 347, and 306 citations. As these papers were published before 2010, we have also selected one most cited paper for 2010-2015 ([4] with 151 citations) and one most cited paper for 2016-2020 ([17] with 78 citations) to validate our comparison framework with more recent studies on IM.

Apply IM Comparison Framework on the Selected Papers. At this step, we characterized the five IM papers with the attributes of the IM Comparison Framework. The details are given in Section 4.

Check IM Comparison Framework Applicability and Completeness. The goal of this step was to verify if each selected IM approach could be characterized using the proposed framework (its *applicability*) and if was complete (its *completeness*).

To find potentially useful papers in scientific bases, we followed the SLR methodology. In IEEE, ACM and Science Direct, we searched till around 625 hits. In Google scholar and Springer, where the search had to be cut into little pieces, 24 queries (corresponding to 24 possible combinations of keywords) were done. In Google scholar, we stopped after 100 hits for each query. In Springer, we checked all retrieved papers. This way of working does not exclude that we missed some interesting papers, especially old ones. That implies our second validity threat: the analysis of several missed papers could have provided some additional details for the IM comparison framework attributes.

3 Comparison Framework

Classification frameworks, such as the one presented in [9-12] are a useful way to introduce a domain and discuss literature in a systematic way. They help characterizing methods techniques and tools, comparing them, demonstrating the originality of given approaches with respect to the rest of the literature, finding gaps in the literature, etc.

As Fig. 2 shows, we propose a classification framework that emphasizes 4 main dimensions of Intention Mining, related to the research questions: the object, the usage, the method, and the tool. More precisely:

- The Object dimension raises the question *What is intention mining?* It refers to the structure of the intentions, their taxonomy and formalism.
- The Usage dimension raises the question *Why using intention mining?* It provides an insight on the different types of objective of intention mining, the domains of application, the source and target of the existing approaches.
- By providing details on intention mining approaches, the Method dimension raises the question *How is intention mining achieved?* For instance, the method dimension characterizes an intention mining approach by indicating the theoretical grounding of the method, whether it exploits classification techniques or ontologies.
- The Tool dimension provides details on support offered to enact the method. The question here is *By which means is the intention mining method put in practice?* Looking at this dimension of an intention mining method indicates the algorithms on which it relies, and the tools that it uses or implements.

The distinction between methods and tools is not obvious; it can be understood with the cooking metaphor: if methods correspond to recipes, tools are kitchen instruments.

Each dimension is specified as a set of attributes with a name and a domain. For instance, the object dimension has 3 attributes: “structuredness” indicates whether the output is an individual intention or a model, “intention taxonomy” and “intention modeling formalism” indicate how mined intentions are further specified, by

classification in taxonomy or by using a specific formalism. Each dimension of the intention mining classification framework is described in a sub-section below.

Intention Mining Comparison Framework			
Object Dimension	Usage Dimension	Method Dimension	Tool Dimension
<ul style="list-style-type: none"> • Structuredness • Intention Taxonomy • Intention Modeling Formalism 	<ul style="list-style-type: none"> • Usage Goal • Domain of application • Source Artefact • Target Artefact 	<ul style="list-style-type: none"> • Machine Learning Method • Automation Method • Observation Record Type • Mathematical Method • Classification-based • Ontology-based 	<ul style="list-style-type: none"> • Mathematical Model • Classification Type • Algorithm Name • Ontology Name • Support • Tool Name

Fig. 2. Intention Mining Comparison Framework.

3.1 The Object Dimension

Table 1 details the object dimension of intention mining that has 3 attributes: *structuredness* defines if the aim is to mine individual intentions (a mining technique can of course still generate collections of individual intentions), or in relationship with other intentions (in which case, the output are intention models); the *intention taxonomy* attribute characterizes the fundamental nature of the mined intentions (several taxonomies can be used in combination, hence the SET OF ENUM type); last the *intention modeling formalism* attribute indicates which notation is used to formalize individual intentions or collections of intentions.

Table 1. Details of the Object Dimension.

Attributes	Values
Structuredness	ENUM (Individual, Model)
Intention Taxonomy	SET OF (ENUM (Action/Semantic; Soft/Hard; Uni-token/Multi-token; Informational/ Ambiguous/Navigational/Transactional; Explicit/Implicit; Objective/Subjective; Research/Purchase; Collective/Individual))
Intention Modeling Formalism	ENUM (Map, I*, Kaos, linguistic, Schemata)

Structuredness. The Object of an IM approach may be to point out intentions either individually, or under the form of structured collections. Each intention can itself be made explicit, e.g. in natural language, or stay implicit as introduced by [18]. For instance, it may be sufficient to find the class of an intention, rather than providing its name or other details [13]. On the other extreme, it is possible to specify intention models, i.e. specifications that conform to a formal notation [6,14]. At the level of collections of intentions, models define dependencies between intentions, such as refinement, abstraction, complementarity, alternatives, positive and negative contributions, etc. At the

individual level, models reveal the underlying structure of intentions, for instance their linguistic structure, their formal condition of realization, etc.

Intention Taxonomy. Several definitions and taxonomies of intentions are proposed in the IM literature. [19] define *action* intentions and *semantic* intentions. The first group represents “basic actions performed on a computer”, the second includes intentions corresponding to “what the user wants to achieve at high level, which may involve several basic actions on a computer to accomplish it”. Whereas *hard* goals can be associated with verification conditions that determine when the goal is achieved, there is no way to determine when *soft* goals are achieved in a clear-cut and definitive way [20]. As an example “protect asset A from any harm” is a typical soft goal: a lot can be done to satisfy it, but whatever is done to protect the asset, nobody can absolutely guarantee that no harm will ever happen to it. [20] presents a taxonomy of web engine user’s intents that emphasizes three main categories: *navigational*, *informational*, and *transactional*. The purpose of a navigational intent is to “reach a particular site”. A web engine user who has an informational intent seeks to “acquire some information assumed to be present on one or more web pages”. Last, the aim of transactional intents is to “perform some web-mediated activity”. This taxonomy is adapted in [21], which generalizes navigational and transactional intents using a concept of non-informational intention, and introduces a concept of ambiguous intention. An even more general taxonomy is the one proposed by [22], which contains 135 high-level intentions grouped in 30 clusters. For [23] an intention is either *objective* when it is about getting factual information, *subjective* when the user wants to collect personal opinions or social when it’s not about getting information but creating interactions with others. For [24], intent is of two commercial types: *research* and *purchase*. [25] consider the taxonomy of intentions based on *implicit* or *explicit* geolocation. [26] considers two other kinds of intentions: *collective* ones and *individual* ones. The collective intentions, often defined as a “commitment of an individual to participate in joint action, and involves an implicit or explicit agreement between the participants to engage in that joint action” are called we-intention, whereas individual intentions are called i-intentions [27].

Intention Modeling Formalism. Besides identifying individual intentions, the object of intention mining approaches may be to generate intentional models specified using a formal notation. Focusing on the IS engineering literature, we distinguish KAOS, *i** and MAP. The KAOS formalism [28] specifies “goals” arranged in a hierarchy, from higher-level goals down to operational goals, i.e. goals that can be operationalized by a system function. Two kinds of dependencies can be specified using KAOS: AND refinement, and OR refinement. The driving principle is that the satisfaction condition of a given higher level goal is semantically equivalent to the conjunction (respectively the disjunction) of the satisfaction condition of the goals that refine it with an AND (respectively OR) refinement link. KAOS models are hierarchy of goals that can be used to elicit, specify, analyze, negotiate, complete, verify, bundle and trace systems requirements. Besides introducing the distinction between “hard goals” and “soft goals”, the *i** framework [29] focuses on the links and dependencies between goals and other concepts such as actors, tasks, resources, etc. and extends the and/or decomposition links with a typology of semantic links such as means/ends, decomposition, dependency, and so-called “contribution links”: makes, breaks, hurts, helps, etc. The MAP formalism [30]

combines the concepts of “intention” and “strategy” in collections of models (called “maps”) organized hierarchically with refinement links. [31] indicates that explicit intentions can be specified with a single word (uni-token) or several words (multi-token). This is further formalized by [32] that uses a linguistic approach inspired by Chomsky’s case grammar to specify the semantic roles of words used in intention names. Another example of individual intention modelling comes from the Schemata method [33] that proposes to model the salient features of individual intentions from a cognitive psychology perspective. Intentions specified with Schemata are for instance mined in [34] to specify intentions in so-called “intention maps” that associate to any web engine query the collection of intentions that motivate it.

3.2 The Usage Dimension

The usage of intention mining approaches can be defined with 4 attributes: the objective of the mined material, the domain of application in which intention mining is used, the source and target users. The domain of these attributes is further defined in Table 2.

Table 2. Details of the Usage Dimension.

Attributes	Values
Usage Goal	ENUM (Discovery, Recommendation, Conformance, Enhancement)
Domain of application	ENUM (Information Retrieval, Robotics, Video, Network forensics, Security, Bioinformatics, Learning, Business Process Modelling, Requirements Engineering, Virtual Consumption, Map Visualization, Game, Database, e-Government, Healthcare, ...)
Source Artefact	ENUM (Individual, Group)
Target Artefact	ENUM (Individual, Group)

Usage Goal. Intentions *discovery* allows understanding people’s intents behind their actions, when executing a query with a search engine [35]. The idea of using intention mining in the context of *recommendation* is that better guidance can be provided if a better knowledge of people’s intentions is acquired [44]. This can be useful to provide personalization, adapted services or products. Conformance and enhancement are two other approaches defined in the Process Mining area [36]. Intention *conformance* aims at controlling whether mined intentions match with their pre-defined specifications. For instance, intention conformance can be used to check that a prescribed model is actually enacted, or measuring the gap between the prescribed models and observing people’s behavior. The idea of *enhancement* is to complete (by refinement, abstraction, or addition) intention models, or improve them in some other way, e.g. with respect to their consistency, feasibility, etc.

Domain of Application. There are many domains of application for IM. Our literature review allowed us to identify: *web engines and information retrieval* [19, 21, 31], which represent the majority of the papers we found, including *online video search* [5, 37]; *network forensics* [38, 39] and *security* [40], *bioinformatics* [41, 42], *business process engineering* [43], *method enactment* [44], *system usage analysis* [45], *learning* [46], *requirements engineering* [32] and *method engineering* [47]. The literature review also included research in *robotics* [16, 48, 49]. Not only we believe this list is not exhaustive, but we have no doubt many more applications will emerge in the future.

Source and Target Artefact. Analysing the aforementioned approaches reveals that the source and target of intention mining can be considered either as *individual* or as a *group*. Intentions can be mined for only one individual (individual sources). In this case, the mined intentions only represent the point of view of one user (obtained from his/her sources). If intentions are mined from many users (group sources), they will represent the point of view of a group of users or “crowd” (obtained from all the logs of the users). The same reasoning can be applied for the target of intention mining. Intentions can be addressed to only one user (individual target) or to many users (group target) depending on the nature of the activity or the objective of the approach.

3.3 The Method Dimension

As shown in table 3, the Method dimension of intention mining can be further refined with the following attributes: Machine learning, to characterize the use of machine learning techniques; Automation method, that indicates the degree of automation used in the method; Observation record type, that indicates what is the input of the method; Mathematical method, that characterizes the theoretical foundation of the mining technique; and Classification-based and Ontology-based, which indicate whether the method exploits a classification technique and ontology.

Table 3. Details of the Method Dimension.

Attributes	Values
Machine Learning	ENUM (Supervised, Unsupervised, Semi-supervised, NULL)
Automation method	ENUM (Manual, Semi-automatic, Automatic)
Observation record type	ENUM (Log, Contextualized trace)
Mathematical method	ENUM (Probabilistic, Statistic, Deterministic, Fuzzy, NULL)
Classification-based	BOOLEAN
Ontology-based	BOOLEAN

Machine Learning. *Supervised* learning is a machine learning technique that consists in inferring a function from labeled training data [50]. *Unsupervised* learning operates on unlabeled data - input where the desired output is unknown [61]. In the context of intention mining, this can be achieved through cluster analysis for instance. *Semi-supervised* learning falls between supervised and unsupervised learning [15].

Automation Method. The automation method attribute characterizes the fact that a method is *manual* (every step of the approach requires a human action) [24], *automatic* (the approach can be fully executed without any intervention) [3, 4, 16] or *semi-automatic* (some steps are manual, others are automatic) [15, 17].

Observation Record Type. Intention mining methods use observations recorded as *logs* or *contextualized traces*. Logs or Traces are temporal sequences of observed events [51]. Existing methods mostly focus on computer logs produced by the users of the systems (as activity, user, timestamp or properties of the used objects) [52]. Enriched log-files contain annotations made by the user while or after performing activities. They are useful in the context of supervised learning methods. Cambridge Dictionary defines a

trace¹ as “a sign that something has happened or existed”, here the focus is on what we expect to observe. Contextual traces record specific actions performed in a specific context; they embed the user activities in the context of their enactment [51]. For instance, software traces can be used to record specific user input, then provided to developers for debugging. [24] use enriched logs containing query, scroll, mouse movements and key press events.

Mathematical Method. Several intention mining methods rely on a mathematical analysis. Their theoretical foundations are: probabilistic, statistic or deterministic. *Deterministic* data do not involve any random variable [53]. When input data are deterministic, then the parameters of the mathematical model are known, and the relationships between variables are strictly functional. In other words, mining techniques based on deterministic models rely on the fundamental assumption that any given input always produces the same output. In *probability* theory, a stochastic process represents an evolution, discrete or continuous, of random variables [53]. Statistical or probabilistic models can be used to deal with stochastic processes. A *statistical* model is a set of probability distribution functions or probability density functions [53]. It involves random variables and assumes that the parameters of a model can be well-estimated using statistical properties of the observed data (as parametric random process). In intention mining approaches that are based on statistical models, the outputs are precisely determined through known statistical relationships between intentions and observations. In the context of intention mining, probabilistic models are analyzed to identify what intentions most probably hide behind the observed behaviors.

Classification and Ontology Based. The added value of ontologies over classifications is that they specify (a) relationships between intentions and other concepts, and (b) reasoning mechanisms.

3.4 The Tool Dimension

Intention mining approaches can be further characterized in terms of the techniques and tools needed to put them into practice. As table 4 shows, the Tool Dimension of the intention mining classification framework has 5 attributes.

Mathematical models. Mathematical models are abstract formal representations used to support reasoning. As we know, two main mathematical models are used in intention mining methods: the Bayesian model [18, 19] and Markov Models [54, 49]. *Bayesian* networks are graphical models that represent random variables and their conditional probabilities via a directed acyclic graph. *Hidden Markov Models* (HMM) are a variant of stochastic Markov chain that represent hidden sequences of states. HMM generalize finite-state automata by evaluating both the probability of transitions between states and probability distributions of observations in those states. The framework *Conditional Random Field* is also used in [24, 55] to define probabilistic models and some works also use other probabilistic models, specially designed for the problem at hand [56, 57].

¹ Cambridge Advanced Learner's Dictionary & Thesaurus <http://dictionary.cambridge.org/dictionary/british/>

Table 4. Details of the Tool Dimension.

Attributes	Values
Mathematical model	ENUM (Bayesian Model, Hidden Markov Model, Conditional Random Field, NULL)
Classification type	ENUM (manual, automatic)
Algorithm name	STRING
Ontology name	STRING
Support	ENUM (sensor, robot, microphone, internet navigation, NULL, etc)
Tool name	STRING

Classification Type. Classification is of first importance in the proposed approaches as the content is often classified to determine the category to which a document or a query belongs. There are a lot of different approaches combining different techniques, like Support Vector Machine [18, 23, 24], Naive Bayes [18, 19], Expectation–maximization [58, 49], Complete-link clustering and Cosine similarity [59]. [14] introduce an improved version of the algorithms proposed by [60], agglomerative clustering using a distance metric based on dynamic time warping (DTW) [61], 1R, J48 and Expectation–Maximization [62]. Sometimes is used Click Intent Model (a hierarchical semantic clustering model) [63], Web Query Classification based on User Intent [64] or even clicks graphs [15].

Algorithm Name. Many algorithms formalize the different stages of IM methods. Among others we can cite for text analysis: TD/IDF [19], Porter Algorithm [31], GBRAM (Goal-Based Requirements Analysis Method) [65], or Markov random walk algorithm and OKAPI BM 25 [3]. Text analysis is one of the first steps of the proposed approaches as queries and document content have to be understandable by algorithms.

Ontology name. Two ontologies are mainly used in intention mining methods: ConceptNet [18, 66] and WordNet [19, 66]. *ConceptNet* [74] is an ontological system for lexical knowledge and common sense knowledge representation and processing. The ConceptNet knowledge base is a semantic network consisting of over 1.6 million assertions encompassing the spatial, physical, social, temporal, and psychological aspects of everyday life. *WordNet* [67] is a lexical database. It groups English words into sets of synonyms, provides short and general definitions, and records various semantic relations between words, such as antonymy, hypernymy and hyponymy. Some ontology-based methods, such as [21], do not use these ontologies, but other ones, like the *Open Directory Project* [68], which categorizes websites. *ASPIC*, a biomedical knowledge base was used in [41] and [69] used the *Library of Congress Subject Headings* as ontology. [61] and others used *Freebase* to create a corpus, so did [3] with *Wikipedia*.

Support. Most of the IM approaches use input logs or user clicks in an Internet navigation [3, 4, 15]. Others approaches include specific devices to support their tool (in [16], a microphone is used to catch human speech to be interpreted by a robot).

Tool Name. Many tools are exploited or proposed to implement intention-mining methods. For instance, [70] and [71] uses the *LIBSVM* tool [72] that implements the Support Vector Machine algorithm. [18] uses the *Natural Language Tool Kit* to manipulate natural language data, and the *WEKA data mining toolkit* for intentions classification. WEKA is widely used for classification [23, 24, 62]. *Mallet* is used by [24] to implement the Conditional Random Field classification algorithm. In [62], a query is transformed into WSMX goals (Web Services Execution Environment) using a system called *Ontopath*. [43]

uses the commercial tool *Disco* [73] to mine processes specified using the BPMN formalism. Of course, some other authors developed a tool from scratch [41,45].

4 Comparison Framework Application

As explained above, we have pre-selected 5 papers to validate the IM comparison framework. [15] use click graphs to better classify users query intents by semi-supervised learning. That allows to avoid to enrich feature representation. [16] aims at detecting affective intents in speech in case of human-robot interaction. Their goal is to learn robots to identify praise, prohibition, comfort, attention, and neutral speech. In [3], the authors present an approach to classify query intents of users by mining the content of Wikipedia. They associate an intent probability to Wikipedia articles and categories to provide a better classification. In the approach described in [4], a new way to capture users' intentions is proposed to improve Internet image search engines. The main idea is to include images in addition to the classic keywords query and to refine search results depending on the clicked image. [17] introduce Capsule Neural Networks to detect user intents when labeled utterances are not available. The proposed approach allows to avoid to label utterances as intents and thus reduce time and labor consumption.

We described the selected IM approaches using the suggested comparison framework (See Table 5). As a result, we were able to check its validity accordingly to two criteria: applicability and completeness. With the first validation criterion, each approach was characterized with regards to the most part of comparison attributes. According to the second criterion, following the application, we found a new attribute - that we called Support - to add to the Tool dimension. Indeed, it is useful to characterize different IM approaches by different kinds of used supports, like specific devices for instance. Based on these findings, we consider that the given comparison framework could be used as a scope for the further systematic literature review. Its application will allow to identify possible values for the framework attributes; thus, it will be possible to provide hints of guidance through the IM approaches selection and to carry out the detailed comparison of the IM approaches based on statistical analysis. In addition, a cross-referenced analysis will allow to identify open issues for future research on IM.

5 Conclusion

This paper proposes a framework to classify intention mining works. Inspired by other classification frameworks, our proposal examines IM in four different views, answering the four main questions about IM called dimensions: "What is IM?", "Why IM?", "How is done IM?", and "By which means is done IM?". Each dimension corresponds to a set of classification attributes, found in the existing literature. We validated our framework by testing it on five existing and good referenced works and concluded that the given framework will be useful to conduct the systematic literature review on IM.

Our next step is to complete our systematic literature review on IM, characterizing all our selected papers within this framework in order to provide a detailed overview of the existing IM works and to define open issues.

Table 5. IM Comparison Framework Application to the Selected IM Approaches..

	[15]	[16]	[3]	[4]	[17]
Object Dimension					
Structuredness	Individual	Individual	Individual	Individual	Cluster
Intention taxonomy	Query intent	“Approval, prohibition, attention, comfort, neutral”	Query intent	Keyword	User intent
Intention Modelling formalism	Linguistic	Linguistic	Linguistic	Linguistic	Linguistic
Usage dimension					
Usage Goal	Discovery of Click Graphs	Discovery of emotional intent	Discovery of category	Recommendation of images	Discovery
Domain of application	Information retrieval	Robot speech	Information retrieval	Internet search image	Information retrieval
Source Artefact	Group	Individual	Group	Individual	Group
Target Artefact	Group	Individual	Individual	Individual	Individual
Method Dimension					
Machine learning Method	Semi-supervised	-	-	-	Semi-supervised
Automation method	Semi-automatic	Automatic	Automatic	Automatic	Semi-automatic (manual labelling)
Observation record type	Query / clickthrough data	Sentence	Search query log / category graph / article graph	Query keywords / query image	Search query log
Mathematical method	-	-	Probabilistic	Probabilistic	Probabilistic
Classification-based	1	1	1	1	1
Ontology-based	0	0	1	0	0
Tool Dimension					
Mathematical model	-	Gaussian model	Markov random walk	Word probability model	-
Classification type	Content based classification (maximum entropy classifier)	EM algorithm / Kurtosis-based approach	Intent classifier	one-class SVM classifier	Intent detection classifier
Algorithm name	“Algorithm 1” / “Algorithm 2”	-	“Algo 1” / “ESA” / “Intent predictor”	-	“Algorithm 1” / “Algorithm 2”
Ontology name	-	-	Wikipedia articles and categories link	-	-
Support	Internet navigation	Microphone	Internet navigation	Internet navigation	Internet navigation

References

1. Merriam-Webster, Definition of Intention, <http://www.merriam-webster.com/dictionary/intention>. Accessed 10 November 2020
2. Bratman, M., Intention, plans, and practical reason. Harvard University Press (1987).
3. Hu, J., Wang, G., Lochovsky, F., Sun, J., Chen, Z.: Understanding user's query intent with wikipedia. In: Proceedings of the 18th international conference on World wide web. pp. 471–480. Association for Computing Machinery, New York, NY, USA (2009).
4. Tang, X., Liu, K., Cui, J., Wen, F., Wang, X.: IntentSearch: Capturing User Intention for One-Click Internet Image Search. *IEEE Transactions on Pattern Analysis and Machine Intelligence*. 34, 1342–1353 (2012).
5. Mei, T., Hua, X.-S., Zhou, H.-Q., Li, S.: Modeling and Mining of Users' Capture Intention for Home Videos. In: *IEEE Transactions on Multimedia*. 9, 66–77 (2007).
6. Khodabandelou, G., Hug, C., Deneckère, R., Salinesi, C.: Supervised intentional process models discovery using Hidden Markov models. In: *IEEE 7th International Conference on Research Challenges in Information Science (RCIS)*. pp. 1–11 (2013).
7. Moshfeghi Y., Joemon J.: On Cognition, Emotion, and Interaction Aspects of Search Tasks with Different Search Intentions. *Proceedings of the 22th International Conference on World Wide Web, WWW 2013*, May, 2013, Rio de Janeiro, Brazil. pp. 931–942. ACM (2009).
8. Dai, H., Zhao, L., Nie, Z., Wen, J.-R., Wang, L., Li, Y.: Detecting online commercial intention (OCI). In: *Proceedings of the 15th international conference on World Wide Web, WWW 2006*, Edinburgh, Scotland, UK, May 23–26, 2006. pp. 829–837. ACM (2006).
9. Jarke, M., Mylopoulos, J., Schmidt, J.M., and Vassilou, Y. (1990) DAIDA : Conceptual Modeling and Knowledge Based Support of Information Systems Development Process. *Technique et Science Informatiques*, 9 (2) : 122 – 133.
10. Rolland, C. A Comprehensive View of Process Engineering. In *10th Conference on Advanced Information Systems Engineering*, Pisa, Italy (1998)
11. Kornysheva, E. and Salinesi, C. MCDM Techniques Selection Approaches: State of the Art. In *IEEE Symposium on Computational Intelligence in Multicriteria Decision Making*, USA (2007)
12. Deneckere, R., Iacovelli, A., Kornysheva, E., and Souveyet, C. From Method Fragments to Method Services. In *Exploring Modeling Methods for Systems Analysis and Design (EMMSAD'08)*, pages pp 80–96, Montpellier, France (2008)
13. Azin, A., Clarke, C., Agichtein, E., Guo, Q.: Estimating Ad Clickthrough Rate through Query Intent Analysis. In: *IEEE/WIC/ACM International Conference on Web Intelligence* (2009)
14. Chapelle, O., Ji, S., Liao, C., Velipasaoglu, E., Lai, L., Wu, S.-L.: Intent-based diversification of web search results: metrics and algorithms. *Inf Retrieval*. 14, 572–592 (2011).
15. Xiao L., Gunawardana, A., Acero, A.: Unsupervised semantic intent discovery from call log acoustics. *IEEE International Conference on Acoustics, Speech, and Signal Processing (ICASSP '05)*, 2005. p. I/45–I/48 Vol. 1 (2005)
16. Breazeal, C., Aryananda, L.: Recognition of Affective Communicative Intent in Robot-Directed Speech. *Autonomous Robots*. 12, 83–104 (2002).
17. Xia, C., Zhang, C., Yan, X., Chang, Y., Yu, P.S.: Zero-shot User Intent Detection via Capsule Neural Networks. Presented at the EMNLP 2018 (2018).
18. Strohmaier M., Kröll M.: Acquiring knowledge about human goals from Search Query Logs. *Information Processing & Management*. 48, 63–82 (2012).
19. Chen, Z., Lin, F., Liu, Y., Ma, W.-Y., Wenyin, L.: User Intention Modeling in Web Applications Using Data Mining. *Conference World Wide Web* (2002).

20. Broder, A.: A taxonomy of web search. *ACM Sigir forum*. pp. 3–10. ACM (2002).
21. Baeza-Yates, R., Calderón-Benavides, L., González-Caro, C.: The Intention Behind Web Queries. In: *String Processing and Information Retrieval*. pp. 98–109. (2006).
22. Chulef, A.S., Read, S.J., Walsh, D.A.: A Hierarchical Taxonomy of Human Goals. *Motivation and Emotion*. 25, 191–232 (2001).
23. Chen, L., Zhang, D., Mark, L.: Understanding user intent in community question answering. *Proceedings of the 21st international conference companion on World Wide Web*. pp. 823–828. ACM (2012).
24. Guo, Q., Agichtein, E.: Ready to buy or just browsing?: detecting web searcher goals from interaction data. *33rd international ACM SIGIR conference on Research and development in information retrieval - SIGIR '10*. p. 130. ACM Press, Geneva, Switzerland (2010).
25. Yi, X., Raghavan, H., Leggetter, C.: Discovering Users' Specific Geo Intention in Web Search. *Proceedings of the 18th International Conference on World Wide Web*. pp. 481–490. ACM, New York, NY, USA (2009).
26. Shen, A.X., Lee, M.K., Cheung, C.M., Chen, H.: An investigation into contribution I-Intention and We-Intention in open web-based encyclopedia: Roles of joint commitment and mutual agreement. *ICIS 2009 Proceedings*. 7 (2009).
27. Tuomela, R.: We-intentions revisited. *Philosophical Studies*. 125, 327–369 (2005).
28. Dardenne, A., van Lamsweerde, A., Fickas, S.: Goal-directed requirements acquisition. *Science of Computer Programming*. 20, 3–50 (1993).
29. Yu, E.S.-K.: *Modelling Strategic Relationships for Process Reengineering*, (1996).
30. Rolland, C., Prakash, N., Benjamin, A.: A Multi-Model View of Process Modelling. *Requirements Eng.* 4, 169–187 (1999).
31. Hashemi, R., Bahrami, A., LaPlant, J., Thurber, K.: Discovery of Intent through the Analysis of Visited Sites. *2008 International Conference on Information & Knowledge Engineering*. pp. 417–422. CSREA Press, Las Vegas, Nevada (2008).
32. Prat, N.: Goal Formalisation and Classification for Requirements Engineering. *Proceedings the 3rd International Workshop on Requirements Engineering*. (1997)
33. Rumelhart, D.E.: *Schemata: The Building Blocks of Cognition*. Lawrence Erlbaum, Hillsdale, NJ (1980).
34. Park, K., Lee, T., Jung, S., Lim, H., Nam, S.: Extracting Search Intentions from Web Search Logs. *2010 2nd International Conference on Information Technology Convergence and Services (ITCS)*. pp. 1–6 (2010).
35. Kaabi, R.S., Souveyet, C.: Capturing Intentional Services with Business Process Maps. *Proceedings of the 1st International conference in Research Challenges in Information Science*. (2007).
36. van der Aalst, W.: *Process Mining*. Springer, Berlin, Heidelberg (2011).
37. Hanjalic, A., Kofler, C., Larson, M.: Intent and its discontents: the user at the wheel of the online video search engine. *20th ACM international conference on Multimedia*. pp. 1239–1248. Association for Computing Machinery, New York, NY, USA (2012).
38. Jantan, A., Rasmi, M., Ibrahim, M.I., Rahman, A.H.A.: A Similarity Model to Estimate Attack Strategy Based on Intentions Analysis for Network Forensics. In: *Recent Trends in Computer Networks and Distributed Systems Security*. pp. 336–346. Springer Berlin Heidelberg (2012).
39. Rasmi, M., Jantan, A.: Attack Intention Analysis Model for Network Forensics. In: *Software Engineering and Computer Systems*. pp. 403–411. Springer Berlin Heidelberg (2011)
40. Shirley, J., Evans, D.: The user is not the enemy: fighting malware by tracking user intentions. *Proceedings of the 2008 workshop on New Security Paradigms*. pp. 33–45. ACM (2009).
41. Sutherland, K., McLeod, K., Ferguson, G., Burger, A.: Knowledge-driven enhancements for task composition in bioinformatics. *BMC Bioinformatics*. 10, S12 (2009).

42. Shanechi, M.M., Wornell, G.W., Williams, Z., Brown, E.N.: A parallel point-process filter for estimation of goal-directed movements from neural signals. *Acoustics Speech and Signal Processing (ICASSP)*, 2010 IEEE International Conference on. pp. 521–524. IEEE (2010).
43. Outmazgin, N., Soffer, P.: Business Process Workarounds: What Can and Cannot Be Detected by Process Mining. In: *Enterprise, Business-Process and Information Systems Modeling*. pp. 48–62. Springer Berlin Heidelberg (2013).
44. Epure, E.V., Hug, C., Deneckere, R., Brinkkemper, S., others: What Shall I Do Next? Intention Mining for Flexible Process Enactment. *Proceedings of 26th International Conference on Advanced Information Systems Engineering (CAiSE)*. pp. 473–487 (2014).
45. Krüger, F., Yordanova, K., Kirste, T.: Tool Support for Probabilistic Intention Recognition Using Plan Synthesis. In: Paternò, F., Ruyter, B. de, Markopoulos, P., Santoro, C., Loenen, E. van, and Luyten, K. (eds.) *Ambient Intelligence*. pp. 439–444. Springer Berlin Heidelberg (2012).
46. Niekum, Scott, Barto, Andrew G.: Clustering via Dirichlet Process Mixture Models for Portable Skill Discovery. *Advances in Neural Information Processing Systems*. Weinberger (2011).
47. Deneckère, R., Hug, C., Khodabandelou, G., Salinesi, C.: Intentional Process Mining: Discovering and Modeling the Goals Behind Processes using Supervised Learning. *International Journal of Information System Modeling and Design*. 5, 22–47 (2014).
48. Aarno, D., Kragic, D.: Layered HMM for Motion Intention Recognition. *2006 IEEE/RSJ International Conference on Intelligent Robots and Systems, IROS 2006*, October 9–15, 2006, Beijing, China. pp. 5130–5135.
49. Bascetta, L., Ferretti, G., Rocco, P., Ardö, H., Bruyninckx, H., Demeester, E., Lello, E.D.: Towards safe human-robot interaction in robotic cells: An approach based on visual tracking and intention estimation. *2011 IEEE/RSJ International Conference on Intelligent Robots and Systems, IROS 2011*, San Francisco, CA, USA, September, 2011. pp. 2971–2978. IEEE (2011).
50. Mohri, M., Rostamizadeh, A., Talwalkar, A.: *Foundations of Machine Learning*. The MIT Press (2012).
51. Laflaquière, J., Settouti, L.S., Prié, Y., Mille, A.: Trace-Based Framework for Experience Management and Engineering. In: *Knowledge-Based Intelligent Information and Engineering Systems*. pp. 1171–1178. Springer Berlin Heidelberg (2006).
52. Clauzel, D., Sehaba, K., Prie, Y.: Modelling and Visualising Traces for Reflexivity in Synchronous Collaborative Systems. *International Conference on Intelligent Networking and Collaborative Systems*, 2009. INCOS '09. pp. 16–23 (2009).
53. Milton, J.S., Arnold, J.C.: *Introduction to Probability and Statistics: Principles and Applications for Engineering and the Computing Sciences*. McGraw-Hill, Inc., New York, NY, USA (2002).
54. Sadikov, E., Madhavan, J., Wang, L., Halevy, A.Y.: Clustering query refinements by user intent. In *Proceedings of the 19th International Conference on World Wide Web, WWW 2010*, Raleigh, North Carolina, USA, April 26–30, 2010. pp. 841–850. ACM (2010).
55. Shen, Y., Yan, J., Yan, S., Ji, L., Liu, N., Chen, Z.: Sparse hidden-dynamics conditional random fields for user intent understanding. In *Proceedings of the 20th International Conference on World Wide Web, WWW 2011*, Hyderabad, India, March 28 - April 1, 2011. pp. 7–16. ACM (2011).
56. Liu, N., Yan, J., Fan, W., Yang, Q., Chen, Z.: Identifying vertical search intention of query through social tagging propagation. *18th International Conference on World Wide Web, WWW 2009*, Madrid, Spain, April 20–24, 2009. pp. 1209–1210. ACM (2009).
57. Xi, X., Wang, Y.-Y., Acero, A.: Learning query intent from regularized click graphs. In: *SIGIR conference on Research and development in information retrieval* (2008).
58. Zhu, C., Cheng, Q., Sheng, W.: Human intention recognition in smart assisted living systems using a hierarchical hidden markov model. *Automation Science and Engineering, CASE 2008*. IEEE International Conference on. pp. 253–258. IEEE (2008).

59. Yamamoto, T., Sakai, T., Iwata, M., Yu, C., Wen, J.-R., Tanaka, K.: The wisdom of advertisers: mining subgoals via query clustering. *Proceedings of the 21st ACM international conference on Information and knowledge management*. pp. 505–514. ACM (2012).
60. Agrawal, R., Gollapudi, S., Halverson, A., Ieong, S.: Diversifying Search Results. *Proceedings of the Second ACM International Conference on Web Search and Data Mining*. pp. 5–14. ACM, New York, NY, USA (2009).
61. Cheung, J.C.K., Li, X.: Sequence clustering and labeling for unsupervised query intent discovery. In: *Proceedings of the fifth ACM international conference on Web search and data mining*. p. ages 383-392 (2012).
62. Ciesielski V., Lalani A.: *Data Mining of Web Access Logs From an Academic Web Site. Design and application of hybrid intelligent systems*. pp. 1034–1043. IOS Press, Amsterdam, The Netherlands (2003).
63. Hakkani-Tür, D., Celikyilmaz, A., Heck, L., Tur, G.: A Weakly-Supervised Approach for Discovering New User Intents from Search Query Logs. (2013).
64. Jansen, B.J., Booth, D.L., Spink, A.: Determining the informational, navigational, and transactional intent of Web queries. *Information Processing & Management*. 44, 1251–1266 (2008).
65. Anton, A.I., Earp, J.B., Reese, A.: Analyzing Website privacy requirements using a privacy goal taxonomy. *IEEE Joint International Conference on Requirements Engineering*, 2002. pp. 23–31 (2002).
66. Krenge, J., Petrushyna, Z., Kravcik, M., Klamma, R.: Identification of learning goals in forum-based communities. *11th IEEE International Conference on Advanced Learning Technologies (ICALT)*. pp. 307–309. IEEE (2011).
67. Miller, G.A., Beckwith, R., Fellbaum, C., Gross, D., Miller, K.J.: Introduction to WordNet: An On-line Lexical Database*. *Int J Lexicography*. 3, 235–244 (1990).
68. DMOZ, <http://www.dmoz.org/> Accessed in November 2020
69. Shen, Y., Li, Y., Xu, Y., Iannella, R., Algarni, A.: An Ontology-based Mining Approach for User Search Intent Discovery. 9 (2011).
70. González-Caro, C., Baeza-Yates, R.: A Multi-faceted Approach to Query Intent Classification. In: Grossi, R., Sebastiani, F., and Silvestri, F. (eds.) *String Processing and Information Retrieval*. pp. 368–379. Springer, Berlin, Heidelberg (2011).
71. Herrera, M.R., de Moura, E.S., Cristo, M., Silva, T.P., da Silva, A.S.: Exploring features for the automatic identification of user goals in web search. *Information processing & management*. 46, 131–142 (2010).
72. Chang, C.-C., Lin, C.-J.: LIBSVM: A Library for Support Vector Machines. *ACM Trans. Intell. Syst. Technol.* 2, 27:1–27:27 (2011).
73. Fluxicon, <http://www.fluxicon.com/disco/> Accessed in November 2020
74. Liu, H., Singh, P.: ConceptNet & Mdash; A Practical Commonsense Reasoning Tool-Kit. *BT Technology Journal*. 22, 211–226 (2004).