Using FCA and Concept Explications for Finding an Appropriate Concept

Marek Menšík[1], Adam Albert[1], and Tomáš Michalovský[2]

[1] VSB - Technical University of Ostrava, Czech Republic
mensikm@gmail.com, adam.albert@vsb.cz,
[2] Silesian University in Opava, Czech Republic
mic0129@slu.cz

Abstract. This paper introduces the method of discovering a plausible atomic concept that corresponds to the generated molecular concept explication and known attributes' values and properties of objects falling under the concept. First, we summarize the process of concept explication via the symbolic method of supervised machine learning from formalized natural language sentences. To obtain particular concept explications, we exploit heuristic procedures that operate on the symbolic representation of current hypothesis and example to obtain particular concept explications. These explications serve as descriptions of the sought atomic concept accordingly to the given text sources. Afterwards, the method of searching for the appropriate concept based on attributes' values is outlined. Thus, user can seek a specific concept, which can be vague or inaccurate, among the so-extracted explications. We focus on a situation in which the user knows basic properties or attributes' values and searches for a suitable atomic concept that is described by these properties or attributes' values. To explain the process, we summarize the creation of explications and the method of Formal Concept Analysis (FCA) as a theoretical background. As a result, we present to the user an appropriate atomic concept. The whole method is demonstrated by a few examples.

Keywords: FCA · NLP · Explications · Formal concept

1 Introduction

The paper is follow up to our current natural language processing research. In [1], we exploited the supervised machine learning for creation of hypothesis that classifies objects. In [2], we modified the algorithm of machine learning for concept refinement in the form of explications obtained from texts in natural language. In [3], the method of seeking appropriate text sources was presented.

In this paper, we deal with the method of recommending an appropriate concept by a given specific set of properties or values of attributes of objects that are falling under the concept. To this date, we have dealt with creating of explications and with the recommendation of a relevant text source based on a chosen explication. In this paper, we decided to reverse this process and by

A. Horák, P. Rychlý, A. Rambousek (eds.): Proceedings of Recent Advances in Slavonic Natural Language Processing, RASLAN 2021, pp. 49–60, 2021.

exploiting FCA we seek a concept that corresponds to the given set. To introduce the reader to this problem, the explication is explained in the beginning.

The explication [4] is a process of refining of an inaccurate or vague expression into an adequately accurate one. For a sake of simplicity, we will refer to refinement as the concept explication.

In prior papers, we have focused on creating of explications of concepts using symbolic methods of supervised machine learning, that utilizes induction heuristics. These functions manipulates with a symbolic representation of explication in process of learning. As for symbolic representation we chose the strong expressiveness of Transparent intensional logic (TIL) and it's computational variant the TIL-script language. TIL and TIL-script are thoroughly described in publications such as [5], [6]. For this reason we will not explain them but we will highlight their features we exploited.

This paper is structured as follows. In chapters 2, the process of creating of explications is described. Chapter 3 summarizes the theory of FCA needed for understanding the *aspirant ordering* used for ordering of concepts by relevancy to the user. In chapter 4 we present the whole process of finding *appropriate concepts* on an example and in last chapter 5 concludes our research.

2 Supervised Machine Learning

Supervised machine learning is a method in which an agent is being trained by classified training examples provided by the supervisor. Examples are described by attributes divided into two groups, namely input and output attributes. There is a functional dependency f between values of those two groups. For example, conditions for receiving a loan by bank can be described by input attributes *employment*, *salary*, *age*, *indebtedness* and *health condition* of an applicant. The risk of providing a loan to the applicant is the output attribute. The goal of the supervised machine learning is that the agent creates his own functional dependency h by observing values of input and output attributes. Agent's functional dependency h, called a *hypothesis*, should approximate the original unknown function f.

Correctness of the learned hypothesis is verified by special set of examples called test examples. The agent knows only the values of input attributes of the test examples. If the hypothesis predicts the same values of output attributes as the original dependency f, the hypothesis is correct. More about supervised machine learning can be found in [7], [8], [9].

2.1 Algorithm framework

As one of the symbolic methods of supervised machine learning, our algorithm can be described by its general framework [8]. This framework consists of four parts: objectives, training data, data representation, and a module that operates on the symbolic representation.

Our adjusted algorithm does not produce hypothesis which would correctly classify unknown examples; it rather builds an explication of an atomic concept C. In TIL, the atomic concept is a Trivialisation of an object X, in symbols, $'X$.[3] The objective of our algorithm is to create the explication an explication in the form of a closed molecular construction producing an object $'Y$ as close to object X as possible.

The natural language sentences mentioning the atomic concept C play the role of training data. Sentences satisfying this condition are formalized into the language of TIL constructions, which serve as a data representation for our algorithm.

The module for manipulation with the symbolic representation contains heuristic functions. For our purpose, we have chosen functions from Patrick Winston's algorithm [10] adjusted for natural language processing. They are divided into two categories of functions.

Functions from the *Generalization* category replace one or more constituents of the hypothesis by a more general one. New or adjusted constituent is either created based on agent's internal ontology or it is created as a disjunction of new and existing constituents. In case of numerical values in the existing constituent and example, generalization can create an interval spanning both numerical values or it can alter existing numerical interval to cover the new value in example. For example, if we have a piece of information in the hypothesis that lions can live up to 10 years on average in the wild and in the example, we have another piece of information that lions can live up to 14 years on average in the wild, generalization will adjust the information in the hypothesis that lions can live up from 10 to 14 years on average in the wild. Thus, our hypothesis becomes more general.

Specialization is triggered by negative examples. In this case, new constituent is inserted into the molecular hypothesis. The constituent doesn't belong into the essence of an explicated object but it helps to distinguish the hypothesis from other similar explications. For instance, the explication of lioness can be specialized with a constituent meaning that lioness does not have a mane. With this information, we can differentiate the explication of lioness from for example an explication of lion.

The original Winston's algorithm [10] deals with examples that cover all the attributes of a learned object. It was not suitable for processing natural-language texts. Sentences that mention explicated object usually do not contain all requisites or typical properties of the object. Since we need to insert new constituents into the explication, we introduced in [2] a new algorithm method called *Refinement*, which contains a single heuristic function for adding a new requisites or typical properties into the hypothesis. More about heuristic functions contained in the generalization, specialization and refinement and about the process of creating explications can be found in [3].

[3] Trivialization $'X$ can be found in other papers written as ^0X.

2.2 Example of Generating an Explication

The symbolic methods of supervised machine learning that we discussed in [1] use heuristic functions which manipulate with a symbolic representation of the hypothesis to obtain the correct one. The language of TIL constructions was chosen for the symbolic representation of hypothesis and examples. This method was then adjusted in [2] for the purpose of the explication of an atomic concept by extracting sentences in natural language texts mentioning the atomic concept as positive and negative examples. Input attributes were in the form of molecular concepts explicating the learnt concept. Output attribute was the atomic concept to be learned. For example, to explicate a atomic concept *lioness*, i.e. Trivialization *'Lioness* of the property of being a lioness of type $(o\iota)_{\tau\omega}$, we can use sentences in natural language which explicate the property. For example, the positive example *"Lioness is a mammal which is an apex predator"*. The property of *being a mammal and apex predator* is formalized in TIL as the following construction.

$$\lambda w \lambda t\, \lambda x\, [[{}'Mammal_{wt}\; x] \wedge [[{}'Apex\; {}'Predator]_{wt}\; x]]$$

Types: $Apex/((o\iota)_{\tau\omega}(o\iota)_{\tau\omega})$: property modifier;
$Predator, Mammal/(o\iota)_{\tau\omega}$: properties of individuals; $w \to \omega$; $t \to \tau$; $x \to \iota$: variables ranging over possible worlds, times and individuals, respectively. TIL and its utilization in the process of explication is described in detail in [2], [3]. Reader can find more about TIL itself in [5], [12].

Training data for our method are natural language sentences. Only sentences mentioning the atomic concept are extracted and formalized into TIL constructions. Agent's hypothesis is refined or generalized by exploiting positive examples. By refinement, we insert new constituents into the hypothesis. With generalization, we adjust current constituents to prevent over specialization of the explication. By negative examples, we specialize the hypothesis to differentiate it from other similar concepts. For example, we can refine the above mentioned explication mentioned above with a positive example in the form of the sentence *"The lioness has a fur."*. The property of *having a fur* formalized in TIL construction as

$$\lambda w \lambda t\, \lambda x\, [{}'Has\text{-}fur_{wt}\; x]$$

Types: $Has\text{-}fur/(o\iota)_{\tau\omega}$: property of individuals.

This positive example triggers a heuristic function that enriches the hypothesis with a new constituent in conjunctive way.

$$\lambda w \lambda t\, \lambda x\, [[{}'Mammal_{wt}\; x] \wedge [[{}'Apex\; {}'Predator]_{wt}\; x] \wedge [{}'Has\text{-}fur_{wt}\; x]]$$

As mentioned above, by generalization, we can avoid having the explication too specific. For example, the explication contains an information that lioness lives in Africa.

$$\lambda w \lambda t\ \lambda x[['Mammal_{wt}\ x] \wedge [['Apex\ 'Predator]_{wt}\ x]$$
$$\wedge\ ['Has\text{-}fur_{wt}\ x] \wedge ['Lives\text{-}in_{wt}\ x\ \lambda w \lambda t\ \lambda y['Africa_{wt}\ y]]]$$

Types: $Lives\text{-}in/(o\iota(o\iota)_{\tau\omega})_{\tau\omega}$; $Africa/(o\iota)_{\tau\omega}$

The explication can be generalized by a positive example *"Lioness lives in India."*. Generalization will adjust existing constituent in disjunctive way, thus making the explication more general.

$$\lambda w \lambda t\ \lambda x[['Mammal_{wt}\ x] \wedge [['Apex\ 'Predator]_{wt}\ x]$$
$$\wedge\ ['Has\text{-}fur_{wt}\ x] \wedge [['Lives\text{-}in_{wt}\ x\ \lambda w \lambda t\ \lambda y['Africa_{wt}\ y]]$$
$$\vee\ ['Lives\text{-}in_{wt}\ x\ \lambda w \lambda t\ \lambda y['India_{wt}\ y]]]]$$

Types: $India/(o\iota)_{\tau\omega}$

3 FCA and Aspirant Ordering

As stated above, the user selects the set of properties and attributes' that should characterise the sought concept. To this end, we exploit the FCA theory that is described in this chapter. The FCA is utilized to obtain all formal concepts and create conceptual lattice over explications.[4] The lattice provides overview of explication ordering. Base on the set of formal concepts we find all *'concept aspirants'*. Concept Aspirants (CA) is the set union of all concepts' intents of which the selected set of properties is an intents' subset. Next the set is ordered and the maximal element of the set is presented to the user as the *most appropriate* one.

As we mentioned in [6]. Formal Conceptual Analysis (FCA) was introduced in 1980s by the group lead by Rudolf Wille and became a popular technique within the information retrieval field.[5] FCA has been applied in many disciplines such as software engineering, machine learning, knowledge discovery and ontology construction. Informally, FCA studies how objects can be hierarchically grouped together with their mutual common attributes.

The following part deals with formal definitions and examples describing the process of selecting the most appropriate concept.

Definition 1. *Let (G, M, I) be a formal context, then $\beta(G, M, I) = \{(O, A) | O \subseteq G, A \subseteq M, A^{\downarrow} = O, O^{\uparrow} = A\}$ is a set of all formal concepts of context (G, M, I) where $I \subseteq G \times M, O^{\uparrow} = \{a | \forall o \in O, (o, a) \in I\}, A^{\downarrow} = \{o | \forall a \in A, (o, a) \in I\}$. $A^{\downarrow}$ is called* ***extent*** *of formal concept (O, A) and $O^{\uparrow}$ is called* ***intent*** *of formal concept (O, A).*

Definition 2. ***Concept aspirants*** *of the set of attributes* ***a*** *in $\beta(G, M, I)$ is a set $CA(a) = \bigcup_{i=1}^{n} O_i^a$, where O^a is extent of a concept $(O, A) \neq (G, B), a \subseteq A, B \subseteq M$. Namely,* ***concept aspirants*** *of the set of attributes* ***a*** *is a union of all formal concept extents where* ***a*** *is a subset of a particular formal concepts' intents.*

[4] In this paper we do not visualise the concept lattice as a graph structure.

[5] More in [13].

Definition 3. *Let $CA(a)$ be a set of concept aspirants of a set of attributes* $\boldsymbol{a}$, *let* $\delta(a)$ *be a set of concepts* (O, A) *where* $a \subseteq A$, *i.e.:* $\delta(a) = \{(O^a, (O^a)^\uparrow) | (O^a, (O^a)^\uparrow) \neq (G, B), B \subseteq M, (O^a, (O^a)^\uparrow) \in \beta(G, M, I)\}$. *Then* $\boldsymbol{x} \sqsubseteq \boldsymbol{y}$ *is in relation of* ***aspirant ordering*** *iff* $max(|(O^y)^\uparrow|) \leq max(|(O^x)^\uparrow|), x, y, \in CA(a), (O^x, (O^x)^\uparrow), (O^y, (O^y)^\uparrow) \in \delta(a)$.

Definition 4. *Let* $(CA(a), \sqsubseteq)$ *be an ordered set according to the definition 3, then the maximal elements are the* ***most appropriate concepts****.*

Example: Let us have a formal context described by the following Table 1 and assume that the user seeks a concept which is described by the set of attributes $a = \{a_2\}$.

Table 1: Formal context

	a_0	a_1	a_2	a_3
o_0	1	1	0	0
o_1	0	1	1	0
o_2	0	1	1	1

The set of all *formal concepts*

$$(G, M, I) = \{C_0, C_1, C_2, C_3, C_4\},$$

where

$$\begin{array}{ll} C_0 = (\{o_0, o_1, o_2\}, \{a_1\}) & C_1 = (\{o_0\}, \{a_0, a_1\}) \\ C_2 = (\{o_1, o_2\}, \{a_1, a_2\}) & C_3 = (\{o_2\}, \{a_1, a_2, a_3\}) \\ C_4 = (\emptyset, \{a_0, a_1, a_2, a_3\}) & \end{array}$$

Find the set of *concept aspirants* for attributes *a*

$$a = \{a_2\}$$

1. Find set $\delta(a)$:
   $\delta(a) = \{(\{o_1, o_2\}, \{a_1, a_2\}), (\{o_2\}, \{a_1, a_2, a_3\}),$
   $(\emptyset, \{a_0, a_1, a_2, a_3\})\}$
2. Create the union of all extents found in step 1 : $CA(\{a_2\}) = \{o_1, o_2\}$
3. For all $x \in CA(\{a_2\})$ calculate max of $|(O^x)^\uparrow|$, where $((O^x), (O^x)^\uparrow) \in \delta(a) \rightarrow max(|\{o_1, o_2\}^\uparrow|) = 2, max(\{|\{o_1, o_2\}^\uparrow|, |o_2{}^\uparrow|\}) = 3$
4. Order $CA(\{a_2\})$ by definition 3 $\rightarrow o_2 \sqsubseteq o_1$

Table 2: Aspirants' ordering

Exp.	Intent	DF
o_1	$\{a_1, a_2\}$	$\{a_1\}$
o_2	$\{a_1.a_2, a_3\}$	$\{a_1, a_3\}$

In our table the column DF represents the difference from the selected set of attributes $\{a_2\}$. The maximum entities according to the orderings are representatives of the most general formal concepts. The most appropriate concept is o_1.

4 Data-set of our Case Study

In this chapter, we present specific explications obtained from several text sources by the algorithm described in chapter 2. Presented explications deal with several concepts of feline predators. These explications are particular samples of all possible explications we can obtain from textual data, because we can obtain several explications of the same concept from different sources. For example, one explication can describe a lioness from an anatomical perspective, another resource may describe the environment in which lioness lives, and still another document describes its behaviour.

The advantage of using the expressive apparatus of TIL is obvious here, since the analyses of sentences that mention the explicated concept are so fine-grained that they are easy to read and understand. Thus, users can easily analyse the differences between particular molecular concepts explicating the target concept. For instance if there are some inconsistencies between the so-obtained explications, the user may exclude those that are not acceptable for him/her. Thanks to this approach, the selection is not based only on syntactic features like the occurrence of a given term, but also on semantic features provided by the fine-grained analysis.

Explications are built up by applying the relation *Typ-p* and the relation *Req* of type $(o(o\iota)_{\tau\omega})(o\iota)_{\tau\omega})$. *Typ-p* is the relation between properties P and Q such that *typically*, if an individual happens to be a Q then most probably it has the property P. For example, the property of *living in Africa* is a typical property of the property of *being a lioness*. On the other hand, *Req*(*uisite*) is a *necessary* relation between properties. Necessarily, if an individual happens to be a lioness, then it must be a mammal as well.

In our example we had at our disposal six explications of atomic concepts, namely explications describing the concepts of 'House cat', 'Jungle cat', 'Sand cat', 'Lynx', 'Lion' and 'Tiger'. All these explications were generated from various sentences formalized into the TIL constructions.

Selected sentences describing the concept 'House Cat':

The house cat is a mammal. The house cat has a fur. The house cat is domesticated. The average height of the house cat is 30cm.

$House\_Cat =$

$$[['Req\ 'Mammal\ ['House\ 'Cat]] \wedge ['Req\ 'Has\text{-}fur\ ['House\ 'Cat]]$$
$$\wedge ['Req\ 'Domesticated\ ['House\ 'Cat]]$$
$$\wedge ['Typ\text{-}p\ \lambda w \lambda t\ \lambda x['= [['Avg\ 'Height]_{wt}\ x]\ '30]\ ['House\ 'Cat]]]$$

The jungle cat is a mammal. The jungle cat has a fur. The average body length of the jungle cat is from 55 to 112 cm. The average height of the jungle cat is 36,5 cm. The fur color of the jungle cat is brown.

$Jungle\_Cat =$

$$[['Req\ 'Mammal\ ['Jungle\ 'Cat]] \wedge ['Req\ 'Has\text{-}fur\ ['Jungle\ 'Cat]]$$
$$\wedge ['Typ\text{-}p\ \lambda w \lambda t\ \lambda x[['\leq ['Bd\text{-}lgth_{wt}\ x]\ '112]$$
$$\wedge\ ['\geq ['Bd\text{-}lght_{wt}\ x]\ '55]]\ ['Jungle\ 'Cat]]$$
$$\wedge ['Typ\text{-}p\ \lambda w \lambda t\ \lambda x['= [['Avg\ 'Height]_{wt}\ x]\ '36.5]\ ['Jungle\ 'Cat]]$$
$$\wedge ['Typ\text{-}p\ \lambda w \lambda t\ \lambda x['=_p ['Fur\text{-}color_{wt}\ x]\ 'Brown]\ ['Jungle\ 'Cat]]]$$

The sand cat is a mammal. The sand cat has a fur. The average body length of the sand cat is from 39 to 57 cm. The average height of the sand cat is 27 cm. The fur color of the sand cat is brown.

$Sand\_Cat =$

$$[['Req\ 'Mammal\ ['Sand\ 'Cat]] \wedge ['Req\ 'Has\text{-}fur\ ['Sand\ 'Cat]]$$
$$\wedge ['Typ\text{-}p\ \lambda w \lambda t\ \lambda x[['\leq ['Bd\text{-}lgth_{wt}\ x]\ '57]$$
$$\wedge\ ['\geq ['Bd\text{-}lght_{wt}\ x]\ '39]]\ ['Sand\ 'Cat]]$$
$$\wedge ['Typ\text{-}p\ \lambda w \lambda t\ \lambda x['= [['Avg\ 'Height]_{wt}\ x]\ '27]\ ['Sand\ 'Cat]]$$
$$\wedge ['Typ\text{-}p\ \lambda w \lambda t\ \lambda x['=_p ['Fur\text{-}color_{wt}\ x]\ 'Brown]\ ['Sand\ 'Cat]]]$$

The lynx is a mammal. The lynx has a fur. The body length of the lynx is less than 148 cm. The average height of the lynx is 75 cm. The lynx is the biggest European feline predator.

$Lynx =$

$$[['Req\ 'Mammal\ 'Lynx] \wedge ['Req\ 'Has\text{-}fur\ 'Lynx]$$
$$\wedge ['Typ\text{-}p\ \lambda w \lambda t\ \lambda x['\leq [['Avg\ 'Bd\text{-}lgth]_{wt}\ x]\ '148]\ 'Lynx]$$
$$\wedge ['Typ\text{-}p\ \lambda w \lambda t\ \lambda x['= [['Avg\ 'Height]_{wt}\ x]\ '75]\ 'Lynx]$$
$$\wedge ['Typ\text{-}p\ \ ['Biggest\ \ ['EU\ \ ['Feline\ 'Predator]]]\ \ 'Lynx]]$$

The lion is a mammal. The lion has a fur. The lion has a mane. The body length of the lion is from 170 to 250 cm.

Lion =

$$[['Req\ 'Mammal\ 'Lion] \wedge ['Req\ 'Has\text{-}fur\ 'Lion]$$
$$\wedge\ ['Req\ 'Pantherinae\ 'Lion]$$
$$\wedge\ ['Typ\text{-}p\ 'Has\text{-}mane\ \ 'Lion] \wedge ['Req\ ['Significant\ 'Sex\text{-}Dimorph]\ 'Lion]$$
$$\wedge\ ['Typ\text{-}p\ \lambda w \lambda t\ \lambda x[['\leq\ ['Bd\text{-}lgth_{wt}\ x]\ '250]$$
$$\wedge\ \ ['\geq\ ['Bd\text{-}lght_{wt}\ x]\ '170]]\ 'Lion]]$$

The tiger is a mammal. The tiger has a fur. The tiger is an apex predator. The average height of the tiger is 117 cm.

Tiger =

$$[['Req\ 'Mammal\ 'Tiger] \wedge ['Req\ 'Has\text{-}fur\ 'Tiger]$$
$$\wedge\ ['Req\ 'Pantherinae\ 'Tiger] \wedge ['Typ\text{-}p\ ['Apex\ 'Predator]\ 'Tiger]$$
$$\wedge\ ['Typ\text{-}p\ \lambda w \lambda t\ \lambda x['=\ [['Avg\ 'Height]_{wt}\ x]\ '117]\ 'Tiger]]$$

Types:
$Req, Typ\text{-}p / (o(o\iota)_{\tau\omega}(o\iota)_{\tau\omega})$, *Bd-Lgth, Height* $/(\tau\iota)_{\tau\omega}$: attributes
Avg $/((\tau\iota)_{\tau\omega}(\tau\iota)_{\tau\omega})$: attribute modifier
Mammal, Cat, Has-Fur, Domesticated, Fur-color, Brown, Lynx, Predator, Lion, Pantherinae, Has-mane, Sex-Dimorph, Tiger $/(o\iota)_{\tau\omega}$: properties
$=_p\ /(o(o\iota)_{\tau\omega}(o\iota)_{\tau\omega})$
$=, \leq, \geq\ /(o\tau\tau)$
Jungle, House, Sand, Feline, EU, Biggest, Apex, Significant $/((o\iota)_{\tau\omega}(o\iota)_{\tau\omega})$: property modifiers
$\wedge / (ooo)$
$x \to \iota$

Seeking the appropriate concept:

At this point we demonstrate the method of dealing with the explications as described in the previous chapter. Having the above introduced explications obtained from natural-language sentences, all the constituents are extracted and arranged into the incidence matrix. Due to lack of space, incidence matrix is represented by transactions in table 3.[6]

Remark: Each object O in table 3 represents one explication of a particular natural language concept. The set of all subconstructions (attributes in table 3) represents *intent* of a particular formal concept. There exists a formal concept $(\{c\}, \{c\}^{\uparrow})$ for each explicated atomic concept c.

Using FCA, all formal concepts were obtained. List of 10 obtained formal concepts is presented in table 4. Due to lack of space in table 4 symbol O represents the set of all objects, i.e. $O = \{JC, SC, HC, Ly, Li, Ti\}$.

All mentioned attributes $A = \{a_1, \ldots, a_{18}\}$ in table 3 represent the following properties in table 5.

[6] More details in [11]

Table 3: Explications and attributes

Explication (O)	Attributes (A)
Jungle cat (JC)	$\{a_1, a_2, a_3, a_4, a_5\}$
Sand cat (SC)	$\{a_1, a_2, a_4, a_6, a_7\}$
House cat (HC)	$\{a_1, a_2, a_8, a_9\}$
Lynx (Ly)	$\{a_1, a_2, a_{10}, a_{11}, a_{12}\}$
Lion (Li)	$\{a_1, a_2, a_{13}, a_{14}, a_{17}, a_{18}\}$
Tiger (Ti)	$\{a_1, a_2, a_{15}, a_{16}, a_{17}\}$

Table 4: Table of all formal concepts

C	Extent	Intent
C_1	O	$\{a_1, a_2\}$
C_2	$\{JC, SC\}$	$\{a_1, a_2, a_4\}$
C_3	$\{Li, Ti\}$	$\{a_1, a_2, a_{17}\}$
C_4	$\{HC\}$	$\{a_1, a_2, a_8, a_9\}$
C_5	$\{JC\}$	$\{a_1, a_2, a_3, a_4, a_5\}$
C_6	$\{SC\}$	$\{a_1, a_2, a_4, a_6, a_7\}$
C_7	$\{Ly\}$	$\{a_1, a_2, a_{10}, a_{11}, a_{12}\}$
C_8	$\{Ti\}$	$\{a_1, a_2, a_{15}, a_{16}, a_{17}\}$
C_9	$\{Li\}$	$\{a_1, a_2, a_{13}, a_{14}, a_{17}, a_{18}\}$
C_{10}	$\emptyset$	A

Assume that the user chooses the attribute a_{17} representing the property of being a *'Pantherinae'* and wants to know, which concept is represented by the chosen attribute most appropriately.

Concept aspirants are found according to definition 2.

$$CA(\{a_{17}\}) = \{Li, Ti\}$$

Afterward, the set $CA(\{a_{17}\})$ is ordered according to definition 3. The final ordering is as follows:

$$Li \sqsubseteq Ti$$

According to definition 4 the entity *Ti* is is a maximal one, and thus the concept of *'being a Tiger'* is presented to the user as the most appropriate one.

5 Conclusion

In this paper, we have described the method of finding an appropriate concept based on properties and attributes' values known by user. The method is based on data mining method of Formal Conceptual Analysis over explications created by the supervised machine learning algorithm. In the beginning, descriptions of concepts, called explications, are created using formalized natural language

Table 5: The list of all properties

a_1	$'Mammal$
a_2	$'Has\text{-}fur$
a_3	$\lambda w \lambda t\, \lambda x[['\leq\ ['Bd\text{-}lgth_{wt}\ x]\ '112] \wedge ['\geq\ ['Bd\text{-}lght_{wt}\ x]\ '55]]$
a_4	$\lambda w \lambda t\, \lambda x['=_p\ ['Fur\text{-}color_{wt}\ x]\ 'Brown]$
a_5	$\lambda w \lambda t\, \lambda x['=\ [['Avg\ 'Height]_{wt}\ x]\ '36.5]$
a_6	$\lambda w \lambda t\, \lambda x[['\leq\ ['Bd\text{-}lgth_{wt}\ x]\ '57] \wedge ['\geq\ ['Bd\text{-}lght_{wt}\ x]\ '39]]$
a_7	$\lambda w \lambda t\, \lambda x['=\ [['Avg\ 'Height]_{wt}\ x]\ '27]$
a_8	$'Domesticated$
a_9	$\lambda w \lambda t\, \lambda x['=\ [['Avg\ 'Height]_{wt}\ x]\ '30]$
a_{10}	$\lambda w \lambda t\, \lambda x['\leq\ [['Avg\ 'Bd\text{-}lgth]_{wt}\ x]\ '148]$
a_{11}	$\lambda w \lambda t\, \lambda x['=\ [['Avg\ 'Height]_{wt}\ x]\ '75]$
a_{12}	$['Biggest\ \ ['EU\ \ ['Feline\ 'Predator]]]$
a_{13}	$'Has\text{-}mane$
a_{14}	$\lambda w \lambda t\, \lambda x[['\leq\ ['Bd\text{-}lgth_{wt}\ x]\ '250] \wedge ['\geq\ ['Bd\text{-}lght_{wt}\ x]\ '170]]$
a_{15}	$['Apex\ 'Predator]$
a_{16}	$\lambda w \lambda t\, \lambda x['=\ [['Avg\ 'Height]_{wt}\ x]\ '117]$
a_{17}	$'Pantherinae$
a_{18}	$['Significant\ 'Sex\text{-}Dimorph]$

sentences by the language of TIL constructions. TIL constructions are inputs for the supervised machine learning algorithm. In the next step, the FCA data mining method is applied on explications to obtain formal concepts. Combining the properties and attribute values provided by the user and with results of FCA, our method offers appropriate concepts which fall under properties and attributes' values provided by the user. The method is demonstrated by an example with 6 explications of different feline predators.

Acknowledgements. This research has been supported by Grant of SGS No. SP2021/87, VSB-Technical University of Ostrava, Czech Republic, "Application of Formal Methods in Knowledge Modelling and Software Engineering IV" and also supported by CZ.02.2.69/0.0/0.0/18_054/0014696 Rozvoj VaV kapacit Slezské univerzity v Opavě and SGS/11/2019 Rozvoj metod teoretické a aplikované informatiky.

References

1. Menšík, M., Duží, M., Albert, A., Patschka, V., Pajr, M., "Machine learning using TIL". In Frontiers in Artificial Intelligence and Applications, Amsterdam: IOS Press, Vol. 321, pp. 344-362, DOI: 10.3233/FAIA200024

2. Menšík, M., Duží, M., Albert, A., Patschka, V., Pajr, M.: ”Refining concepts by machine learning”. In Computación y Sistemas, Vol. 23, No. 3, 2019, pp. 943–958, doi: 10.13053/CyS-23-3-3242
3. Menšík, M., Duží, M., Albert, A., Patschka, V., Pajr, M.: ”Seeking relevant information sources”. In Informatics'2019, IEEE 15th International Scientific Conference on Informatics, Poprad, Slovakia, 2019, pp. 271-276.
4. Carnap, R: Meaning and Necessity: ”A Study in Semantics and Modal Logic”, Chicago, University of Chicago Press, 1964.
5. Duží, M., Jespersen, B.,Materna, P.: ”Procedural Semantics for Hyperintensional Logic”. Foundations and Applications of Transparent Intensional Logic.”, Berlin: Springer, 2010.
6. Menšík, M., Albert, A., Patschka, V.: ”Using FCA for seeking relevant information sources”. In Recent Advances in Slavonic Natural Language Processing, Volume 2020, pp. 47-54, 2020, ISBN 978-802631600-8
7. Russell, S.J.,Norvig, P.: ”Artificial intelligence: a modern approach.”, 2nd ed. Harlow, Pearson Education, 2014. ISBN 978-1-29202-420-2.
8. Mitchell, T.M.: ”Machine learning”, New York: McGraw-Hill, 1997. ISBN 00-704-2807-7.
9. Poole, D.L., Mackworth, A.K.: ”Artificial intelligence: foundations of computational agents. 2nd pub.”, Cambridge: Cambridge University Press, 2010, ISBN 978-0-521-51900-7.
10. Winston, P. H.: ”Artificial Intelligence”. 3rd ed., Mass.: Addison-Wesley Pub. Co., 1992. ISBN 02-015-3377-4.
11. Albert, A., Duží, M., Menšík, M.,Patschka, V.,Pajr, M.: ”Search for appropriate textual information sources”. In Frontiers in Artificial Intelligence and Applications, Amsterdam: IOS Press, Vol. 333, pp. 227-246, DOI: 10.3233/FAIA200832
12. Tichý, P.: ”The Foundations of Frege's Logic.”, Berlin, New York: De Gruyter, 1988.
13. Ganter, B., Wille, R.: ”Formal Concept Analysis: Mathematical Foundations”. 1st ed., Berlin: Springer, 1999, ISBN 978-3-540-62771-5.