Linguistic Alerts in Information Filtering Systems
Towards Technical Implementations of Cognitive Semantics

Radoslaw P. Katarzyniak[1], Wojciech A. Lorkiewicz[1] and Ondrej Krejcar[2]

[1] *Faculty of Computer Science and Management, Wroclaw University of Technology, Wybrzeze Wyspianskiego 27, 50-370 Wroclaw, Poland*

[2] *Center for Basic and Applied Research, Faculty of Informatics and Management, University of Hradec Kralove, Rokitanskeho 62, 500 03 Hradec Kralove, Czech Republic*

Keywords: Information Filtering, Linguistic Alerts, Computational Semiotics, Epistemic Modalities, Cognitive Semantics.

Abstract: An original model of natural language alerts production is proposed. The alerts are produced by information filtering system and stated in a quasi-natural language, both potentially written and vocalized. The alerts are chosen with respect to a certain collection of uncertain decision rules, thus they inherit various levels of epistemic uncertainty. The quasi-natural language statements include linguistic operators of epistemic modality, as their necessary parts. The proposed model implements in a technical context an adequate cognitive semantics captured by an original theory of epistemic modality grounding defined elsewhere.

1 INTRODUCTION

Users' selective dissemination of information and related information filtering (IF for short) are important challenges for modern information systems (Hanani et al., 2001). They seem particularly crucial for management executives, interested in and strongly dependent on up-to date information related to their everyday business activities (Xu et al., 2011). The way how the selected (filtered) information is presented needs to be designed with substantial influence of real environments in which the executives work, including these days frequent mobility of their daily work. In such circumstances all easily comprehensible presentation modes, for instance applications of quasi-natural, written and sometimes even vocalized languages, have become a very important theoretical and practical problem for computer science community.

Unfortunately in actual settings, it is often the case that on-line indexing of documents, incoming to executives' knowledge repositories, is practically impossible due to their inherent characteristics. For instance, a typical document can consists of expanded multimedia elements and therefore require advanced and time-consuming processing to elaborate semantic description of their content. Fortunately, at least in some practical contexts, an approximate (yet still effective) solution is to settle executive-oriented filtering solely on attributes of incoming documents for which values can be easily determined. Such attributes may include origin, author(s), affiliating institutions, attached general keywords, etc. However, a rather obvious inconvenience of the approximate solution is that filtering decisions may be uncertain to some extent. In particular, due to underlying soft classification rules in which preconditions are defined by means of easy-to-determine attributes, and post-conditions are built from subjects (topics) the executives are interested in. Another inconvenience might be that such IF systems need to be based on processes of classification rules management (namely, their effective extraction, storage, retrieval and update).

In this paper we provide a theoretical background for solving the highlighted problem for a particular class of IF systems. Namely, a theoretical foundation for production of incoming documents' alerts, founded on uncertain classification rules, is discussed. An important functional assumption is that ***alerts are to be stated in quasi-natural languages with linguistic markers for communicating levels of epistemic uncertainty***. In perhaps all languages such linguistic markers exist, usually in a form of well-known and widely used basic modal operators of knowledge *(I know that ...)*, believe *(I believe that ...)* and possibility *(I find it possible/it is possible that ...)*, as well as their possible extensions e.g. *I strongly believe*

Katarzyniak, R., Lorkiewicz, W. and Krejcar, O.
Linguistic Alerts in Information Filtering Systems - Towards Technical Implementations of Cognitive Semantics.
In *Proceedings of the 18th International Conference on Enterprise Information Systems (ICEIS 2016) - Volume 1*, pages 512-519
ISBN: 978-989-758-187-8

that ... (Nuyts, 2001). The way, in which the natural language operators of epistemic modality should be chosen and used as components of information filtering alerts, is an original contribution of this work, comparing to other works, usually dealing with other classes of language vagueness e.g. (Herrera-Viedma et al., 2004).

The overall organization of the presentation is as follows. In section 2, our original model of knowledge base and basic knowledge management processes, underlying the extraction and application of classification rules to alerts' generation, is presented. In this section a concept of mental language holons is introduced as a key knowledge structure participating to adequate alert's choice and extraction. According to their definition, the holons cover complementary language oriented experience summarizations. In section 3, the so-called theory of modality grounding, originally proposed elsewhere (Katarzyniak, 2005), is applied to define strong and logically consistent support for choosing adequate epistemic modality of alerts. The theory is based on a technical model of so-called cognitive semantics for natural language statements, limited to some scope of quasi-natural language statements, modal in the epistemic sense. The section consists a brief note on the novelty of our approach, considered in respect to technical application of cognitive linguistics. In section 4 a computational example is presented. Finally, section 5 summarizes presented results and points out possible future extensions.

2 MODEL DEFINITION

2.1 Profile of User Information Needs and Filtering Task

The system's user[1] is focused on a given aspect (important to the user) of information stored within the system. This relevant to the user context strictly depends on current user's preferences and is inherently individual. This user's focus represents a set of topics of interest (document's subjects or themes) that are of special importance or significance to the user.

The user's information needs are represented by a set of subjects (also called *themes*) $S = \{sub_1, sub_2, ..., sub_M\}$, being of potential interest to him or her. Moreover, we further assume that information needs represent the sole information that the system captures about the user. Consequently, user's information needs represent the user's profile stored within the system.

We further assume that all of the incoming documents are processed (filtered and indexed) in order to determine whether they cover any of the highlighted topics of interest (user's information needs). The sole purpose of the filtering stage is to identify documents that are significant to the end user. In particular, the goal is to identify documents d that having a complete knowledge about them (for instance through thorough manual examination by an expert) would lead to formulation of basic statements – *"d is about sub_j"* or *"d is about sub_j and sub_k"* (where j is different from k). Recognizing such documents should further lead to generation of adequate alerts by the system, to inform the end user about the appearance of important documents.

Seemingly the introduced form of user profile is extremely trivial, as compared to apparently more complex models of user profiles studied and applied in the field of IF systems e.g. (Brown and Jones, 2001; Shapira et al., 1999; Xu et al., 2011). However, as it turns out in the context of presented approach to generation of linguistic alerts even in such an oversimplified profile posses significant problems. In particular, linguistic approach requires application of unconventional linguistic semantical models that represent solid and adequate theoretical background for technical implementations of cognitive semantics for such linguistic alerts.

Importantly, we should highlight that the aforementioned task of document filtering is highly complex. In automatic approaches it requires a complex and computationally exhaustive procedure that is able to analyse the content of a document and determine it's relevance against the set of identified subjects. Moreover, in some realities a fully automatic approach might not even be available, as such a set of semi-automatic or even manual methods must be utilised. Furthermore, in systems with strict processing time restrictions the filtering process posses significant technical problems, both methodologically and computationally.

2.2 The Repository Databases

The repository consists of two classes of documents: already stored documents $\mathbf{D} = \{d_1, d_2, ..., d_K\}$ with complete descriptions (including a description of their semantic content) stored in a regular database of the repository, and new documents (new arrivals) $\mathbf{D}^{new} = \{d_1^{new}, d_2^{new}, ..., d_{K_{new}}^{new}\}$ with incomplete descriptions of

[1] Due to strict editorial limitations, we focus solely on a single user case. However, in a more practical realisation the proposed approach can be easily extended to a case of multiple users.

their thematic content, thus awaiting off-line semantic analysis.

Formally the repository sub-databases can be described as an information system by Pawlak (Pawlak and Skowron, 2007), tailored to our practical context. Let $Rep = (\mathbf{D} \cup \mathbf{D}^{new}, A, V, \rho)$ be further considered, where $\mathbf{D}$ and $\mathbf{D}^{new}$ are sets of stored documents and new arrivals, respectively, $A = \{w_1, w_2, ..., w_L, sub_1, sub_2, ..., sub_M\}$ is a set of attributes, $V = \bigcup_{a \in A} V_a$, where $V_{w_i} = W_i$ and $V_{sub_i} = \{\varepsilon, 0, 1\}$, is a set of attributes' values, and $\rho : \mathbf{D} \times A \to V$ is a partial information function.

The partiality of function ρ reflects the extent to which documents are described, regarding their thematic content (their semantics). Namely, it is assumed that $W = \{w_1, w_2, ..., w_L\}$ consists of multivalued attributes, called conditional ones. Values of conditional attributes are usually delivered at document's arrival, as the attributes represent a set of easily computed parameters/characteristics of the document (computed on-line). Contrary, $S = \{sub_1, sub_2, ..., sub_M\}$ consists of attributes, called thematic attributes, representing the content of documents (in respect to a given profile of information needs). Determining the value of thematic attributes requires intensive (both methodological and computational) off-line semantic analysis of the document.

For the sake of clarity and ease of presentation some additional symbols are further introduced.

Namely, for each document $d \in \mathbf{D} \cup \mathbf{D}^{new}$, $\rho_{d|W} : W \to \bigcup_{i=1}^{L}(W_i)$ is *a conditional-part information function related to document d*, such that for each attribute $x \in W$, $\rho_{d|W}(x) \in W_x$ holds, provided that W_x consists of all possible values of x.

Similarly, for each document $d \in \mathbf{D} \cup \mathbf{D}^{new}$, $\rho_{d|S} : S \to \{\varepsilon, 0, 1\}$ is *a thematic-part information function related to document d*. However, in this case rules for assigning attribute values differ for $d \in \mathbf{D}$ and $d \in \mathbf{D}^{new}$. Namely for each attribute $x \in S$ and each document $d \in \mathbf{D}$, $\rho_{d|S}(x) = 1$ if and only if document d is indexed as being about subject x. Otherwise $\rho_{d|S}(x) = 0$. At the same time, for each attribute $x \in S$ and $d \in \mathbf{D}^{new}$, the value of x is treated as unknown, what is formally represented by $\rho_{d|S}(x) = \varepsilon$.

2.3 Mental Language Holons as Representation of Subject Distribution

As aforementioned, the introduced IF system is dedicated to analyse incoming documents, regarding individual subjects $sub \in S$ or/and their conjunctions $sub_x \wedge sub_y$, where $sub_x \in S, sub_y \in S, sub_x \neq sub_y$. Results from this analysis may be uncertain predictions, communicated by the means of natural language operators of epistemic modality.

Below an adequate model of database meta-descriptions used in the filtering process is proposed. It's purpose is to enable effective and semantically valid realization of the assumed functional IF system's goal. The model will be fully compatible with an original theory of epistemic modality grounding, partially presented in (Katarzyniak, 2005; Katarzyniak, 2006b; Katarzyniak, 2006a). The main assumption of the theory is that linguistic alerts are inseparably connected to (in a sense *grounded in*) so-called mental language holons . Language holons represent embedded summarization of empirical episodic experiences, i.e., experiences strictly related to particular subjects or their binary conjunctions. In many ways language holons are similar to mental models, known from the cognitive linguistics and psychology (Johnson-Laird, 1985). For the sake of completeness it is worth mentioning that, at the technical level, mental language holons can be treated as complexes of complementary classification rules.

In order to formally capture the latter, the following three retrieval languages are introduced:

$$\begin{aligned} &\mathcal{KS} = \{sub_1, sub_2, ..., sub_M\}, \\ &\mathcal{KB} = \{sub_x \wedge sub_y \mid sub_x, sub_y \in S \wedge x < y\}, \\ &\mathcal{KL} = \{\bigwedge_{i=1}^{L}(w_i = x_i) \mid w_i \in W, x_i \in W_i, i = 1..L\}. \end{aligned} \tag{1}$$

The semantics of retrieval languages is given by following functions:

$$\begin{aligned} &\delta_{|\mathcal{KS}} : \mathcal{KS} \to 2^{\mathbf{D}}, \\ &\delta_{|\mathcal{KB}} : \mathcal{KB} \to 2^{\mathbf{D}}, \\ &\delta_{|\mathcal{KL}} : \mathcal{KL} \to 2^{\mathbf{D} \cup \mathbf{D}^{new}}. \end{aligned} \tag{2}$$

where:

$$\begin{aligned} \delta_{|\mathcal{KS}}(sub) &= \{d \in \mathbf{D} \mid \rho_{d|S}(sub) = 1\}, \\ \delta_{|\mathcal{KB}}(sub_x \wedge sub_y) &= \delta_{|\mathcal{KS}}(sub_x) \cap \delta_{|\mathcal{KS}}(sub_y), \\ \delta_{|\mathcal{KL}}(\bigwedge_{i=1}^{L}(w_i = x_i)) &= \{d \in \mathbf{D} \mid \bigwedge_{i=1}^{L}(\rho_{d|W}(w_i) = x_i)\} \end{aligned} \tag{3}$$

Mental language holons are defined for simple subjects in $\mathcal{KS}$ and conjunctive subjects in $\mathcal{KB}$, in respect to particular conditions from retrieval language $\mathcal{KL}^+ \subseteq \mathcal{KL}$, where the subset (of non-empty conditions) $\mathcal{KL}^+$ is defined as: $\mathcal{KL}^+ = \{k \in \mathcal{KL} \mid \delta_{|\mathcal{KL}}(k) \cap \mathbf{D} \neq \emptyset\}$.

Having defined $\mathcal{KL}^+$, we can introduce two auxiliary symbols class K_i and class extension $EXT(K_i)$. In particular, a class K_i defines a set of indistinguishable (conditional attribute-wise κ_i) already process documents, whereas class extension $EXT(K_i)$ defines a set of indistinguishable (conditional attribute-wise κ_i) all documents. Namely, if (and only if) $|\mathcal{KL}^+| = Q \geq 1$ and $\mathcal{KL}^+ = \{\kappa_1, \kappa_2, ..., \kappa_Q\}$, then for $i = 1..Q$,

$$\begin{aligned} K_i &= \delta_{|\mathcal{KL}}(\kappa_i) \cap \mathbf{D}, \\ EXT(K_i) &= \delta_{|\mathcal{KL}}(\kappa_i) \cap \mathbf{D}^{new}. \end{aligned} \tag{4}$$

For each $sub \in S$ and $\kappa_i \in \mathcal{KL}^+$, the (simple subject) mental language holon is given as a vector $simholon$:

$$simholon[\kappa_i, sub, \lambda_A^+(sub), \lambda_A^-(sub)], \tag{5}$$

where

$$\begin{aligned} \lambda_A^+(sub) &= \frac{|\delta_{|\mathcal{KS}}(sub) \cap K_i|}{|K_i|}, \\ \lambda_A^-(sub) &= \frac{|(\mathbf{D} \setminus \delta_{|\mathcal{KS}}(sub)) \cap K_i|}{|K_i|}. \end{aligned} \tag{6}$$

For each conjunctive subject $(sub_x \wedge sub_y) = sub_{xy} \in \mathcal{KB}$ and $\kappa_i \in \mathcal{KL}^+$, the (conjunctive subject) mental language holon is given as a vector $conholon$:

$$\begin{aligned} conholon[\kappa_i, sub_{xy}, \lambda_C^{++}(sub_{xy}), \lambda_C^{+-}(sub_{xy}), \\ \lambda_C^{-+}(sub_{xy}), \lambda_C^{--}(sub_{xy})], \end{aligned} \tag{7}$$

where

$$\begin{aligned} \lambda_C^{++}(sub_{xy}) &= \frac{|\delta_{|D}(sub_x) \cap \delta_{|D}(sub_y) \cap K_i|}{|K_i|}, \\ \lambda_C^{+-}(sub_{xy}) &= \frac{|\delta_{|D}(sub_x) \cap (\mathbf{D} \setminus \delta_{|D}(sub_y)) \cap K_i|}{|K_i|}, \\ \lambda_C^{-+}(sub_{xy}) &= \frac{|(\mathbf{D} \setminus \delta_{|D}(sub_x)) \cap \delta_{|D}(sub_y) \cap K_i|}{|K_i|}, \\ \lambda_C^{--}(sub_{xy}) &= \frac{|(\mathbf{D} \setminus \delta_{|D}(sub_x)) \cap (\mathbf{D} \setminus \delta_{|D}(sub_y)) \cap K_i|}{|K_i|}. \end{aligned} \tag{8}$$

From the pragmatic point of view, mental language holons are higher level summarizations (semantic generalizations) of relative share of complementary bodies of experiences, related to particular subjects (or their conjunctions). The whole repository of language holons, available to IF system's processes and, in particularly to alerts production procedures, is given as follows:

$$\begin{aligned} HOLONS =& SIMHOLONS \cup CONHOLONS, \\ SIMHOLONS =& \{simholon[\kappa, x, \lambda_A^+(x), \lambda_A^-(x)] \\ & \mid \kappa \in \mathcal{KL}^+, x \in \mathcal{KS}\}, \\ CONHOLONS =& \{conholon[\kappa, x, \lambda_C^{++}(x), \lambda_C^{+-}(x), \lambda_C^{-+}(x), \\ & \lambda_C^{--}(x)] \mid \kappa \in \mathcal{KL}^+, x \in \mathcal{KB}\}. \end{aligned} \tag{9}$$

3 ALERTS PRODUCTION

3.1 Alerts and their Semantic Proto-forms

Examples of possible structure and content of alerts, considered in our research, are given as follows:

IF SYSTEM ALERT: There is a new [document: x]. I believe it is about [subject: sub]. You may be interested in reading it!

IF SYSTEM ALERT: Documents [documents: $x_1, ..., x_k$] are new. It is possible that they are about [subjects: sub_x and sub_y]. Should I put them on your pending list?

IF SYSTEM ALERT: There is a new [document: x] worth of being looked at. I believe it is about [subject: sub_x], but not about [subject: sub_y]. According to what I know about your interests, the first issue may be of interest to you. Should I put the document to your working box? Please, answer [YES/NO]!

IF SYSTEM ALERT: Among others, the following documents: $x_1, ..., x_k$ have been received from [source: source], too. I believe they are not about [subject:sub] which you pointed at as your main issue. Whether, despite this shall I put them on your pending list? Please, answer [YES/NO]!

IF SYSTEM ALERT: It is possible that the following [incomings: $x_1, ..., x_L$] deal with [subject:sub], which is on your list of interests. Are you interested in reading them before turning them to our central document base? Please, answer [YES/NO]!

The structure of alerts fully depends on designer's choice and, obviously, it should reflect favoured modes and preferences of particular user (users' group) interactions. In our case the alerts are represented (communicated) in a natural language, which is a partially controlled version of actual language. In advanced multimedia systems the alerts can be vocalized, too.

The common feature of the above examples is their underlying sense. Namely, regardless of their form (individual document vs. group of documents, simple subject vs. conjunctive subject), they all are founded on the same propositional aspect: being about or not being about a particular simple subject (or conjunction of simple subjects). Moreover, For $x \in D \cup D^{new}$ and $sub \in \mathcal{KS} \cup \mathcal{KB}$, each example is

originally created as instantiation (concretization) of one of the following basic linguistic proto-forms:
knowing([document(s):x] is about [subject(s):sub])
believing([document(s):x] is about [subject(s):sub])
possible([document(s):x] is about [subject(s):sub])
or another proto-form, complementary to the above enumerated ones.
It is worth of mentioning that for a fixed document x and a fixed subject *sub* (a simple subject or a binary conjunction of simple subjects) one and only one proto-form should be instantiated as proper representation of epistemic state. Namely, such constraint follows from common sense, natural language pragmatics rule, saying that knowing, believing and finding something only as possible (in the epistemic sense) are mutually exclusive, different states of the same mental epistemic attitude. Thus, in our research an adequate extraction of natural language alerts from IF system's knowledge base (or more strictly: proper and adequate choice and further instantiation of proto-form) becomes a fundamental issue to be elaborated, on both technical and theoretical levels.

In conclusion, similarly to other natural language statements, three aspects of alerts need to be taken into account: *propositional element, modality, and temporal frame*. As it has just been mentioned above, the propositional element is given by predication, which on written (or vocalized) level is referred to by elements of sets $\mathcal{KS}$ and $\mathcal{KB}$. The alerts' temporal dimension is quite apparent. Namely, they are stated in the present grammatical time. A more problematic issue is the alerts' modality choice, which in our case should reflect a kind of epistemic uncertainty of IF system, itself. An important question, of both theoretical and technical nature, is how to properly choose adequate modality markers, in order to extend written (or vocalized) representation of predication (applied to incoming documents). This question is strongly supported by an original theory of grounding of modal epistemic statements, briefly presented below.

3.2 Applying the Theory of Epistemic Modality Grounding to Alerts' Production

The decision rules for proper choice of an adequate modal proto-form, its instantiation (and further presentation to an end user in a written and/or vocalized form) follow from an original theory of grounding, presented elsewhere. Namely, for the case of simple subject-based predication the introductory theoretical results can be found in (Katarzyniak, 2005), for binary conjunctive subject-based predication in (Katarzyniak, 2006b; Katarzyniak, 2006a).

It is assumed in the theory (following multiple models of language production (Evans and Green, 2006; Stachowiak, 2013; Wlodarczyk, 2013)) that particular epistemic operators of modality are related to summarized empirical experience, supporting related language proto-forms. However, these proto-forms are never stored and processed as separate entities, for they are conceptually (mentally) related to their complementary counterparts. In particular, such complexes of complementary proto-forms constitute linguistic holons, which in our technical approach are strongly related to the concept of mental language holons, defined in the previous sections. In consequence, to each linguistic proto-form, always related to one and only one part of a relevant mental language holon, certain intensity of summarized (embodied) experience of a subject (or binary conjunctive subject) is assigned. In the theory of grounding this intensity is numerically represented by the relative grounding strength.

According to the theory of simple modalities grounding, the proper choice of adequate linguistic proto-form is possible if and only if a proper system of the so-called modality thresholds is applied (and technically realized in a system). In our case the system needs to consist of two interrelated sub-systems of thresholds $\{\lambda^{KS}_{Know}, \lambda^{KS}_{maxBel}, \lambda^{KS}_{minBel}, \lambda^{KS}_{maxPos}, \lambda^{KS}_{minPos}\}$ and $\{\lambda^{\wedge}_{Know}, \lambda^{\wedge}_{maxBel}, \lambda^{\wedge}_{minBel}, \lambda^{\wedge}_{maxPos}, \lambda^{\wedge}_{minPos}\}$, for effective control of simple-subject predication instantiation and conjunctive subject predication instantiation, respectively.

An interesting result from the theory of grounding, for the practice perhaps the most important one, is that the system of modality thresholds cannot be freely chosen. Namely, in order to guarantee common sense consistency of (written and verbal) language behaviour the system of modality thresholds has to fulfil some predefined set of requirements, accepted in the theory of grounding, as a reflection of common sense pragmatics applied in actual contexts to natural language operators of knowledge, belief, and possibility. The fact that written and/or verbal behaviour, produced by a technical system based on the theory of grounding, is actually consistent, from the semiotic and pragmatic point of view, can be analytically proved and verified[2].

Moreover, within the numerical scope which is permissible according to the theory of grounding, values for thresholds can be chosen in an arbitrary manner (Katarzyniak, 2005). However, for the case of

[2]Some of the results can be found in (Katarzyniak, 2005; Katarzyniak, 2006b; Katarzyniak, 2006a).

populations of artificial agents it is be possible to obtain them from computationally realized processes of artificial language semiosis (Lorkiewicz et al., 2011).

In order to omit deeper discussion of the theory of grounding (outside of the scope of this work) we further present an original application of the theory to basic rules definition for modal alerts' acceptability and adequacy. The fundamental assumption is that a given modal alert can be produced (by IF system) if and only if its underlying linguistic proto-form is well-grounded in IF system's knowledge base. It means, too, that in this practical context, for a certain alert *being well grounded* is equivalent to *adequately describing a related IF system's state of knowledge about possibility of a certain document* $d \in D \cup D^{new}$ *to deal with a certain subject* $sub \in \mathcal{KS} \cup \mathcal{KB}$. In particular, for any document $d \in D^*$, $d \in EXT(K_i)$, and $sub \in \mathcal{KS}$, the following set of so-called grounding relations constitute the theoretical foundation of IF alerting processes:

$simholon[\kappa_i, sub, \lambda_A^+(sub), \lambda_A^-(sub)]$
$\models_G$ **possible**$([d]$ is about $[sub])$
holds if and only if $\lambda^{KS}_{\mathbf{minPos}} \leq \lambda_A^+(sub) < \lambda^{KS}_{\mathbf{maxPos}}$.

$simholon[\kappa_i, sub, \lambda_A^+(sub), \lambda_A^-(sub)]$
$\models_G$ **believing**$([d]$ is about $[sub])$
holds if and only if $\lambda^{KS}_{\mathbf{minBel}} \leq \lambda_A^+(sub) < \lambda^{KS}_{\mathbf{maxBel}}$.

$simholon[\kappa_i, sub, \lambda_A^+(sub), \lambda_A^-(sub)]$
$\models_G$ **knowing**$([d]$ is about $[sub])$
holds if and only if $\lambda_A^+(sub) = \lambda^{KS}_{\mathbf{Know}} \mathbf{= 1}$.

Rather obviously, complementary alerts on *document* $d \in D^*$ *not being about a particular subject* $sub \in \mathcal{KS}$, are produced with respect to the next three definitions:

$simholon[\kappa_i, sub, \lambda_A^+(sub), \lambda_A^-(sub)]$
$\models_G$ **possible**$([d]$ is **not** about $[sub])$
holds if and only if $\lambda^{KS}_{\mathbf{minPos}} \leq \lambda_A^-(sub) < \lambda^{KS}_{\mathbf{maxPos}}$.

$simholon[\kappa_i, sub, \lambda_A^+(sub), \lambda_A^-(sub)]$
$\models_G$ **believing**$([d]$ is **not** about $[sub])$
holds if and only if $\lambda^{KS}_{\mathbf{minBel}} \leq \lambda_A^-(sub) < \lambda^{KS}_{\mathbf{maxBel}}$.

$simholon[\kappa_i, sub, \lambda_A^+(sub), \lambda_A^-(sub)]$
$\models_G$ **knowing**$([d]$ is **not** about $[sub])$
holds if and only if $\lambda_A^-(sub) = \lambda^{KS}_{\mathbf{Know}} \mathbf{= 1}$.

Obviously, similar set of definitions, for $d \in D^*$, $d \in EXT(K_i)$, and $(sub_x \wedge sub_y) = sub_{xy} \in \mathcal{KB}$, can also be formulated and used, if needed. However, in such case another mental language holons must be referred to:

$conholon[\kappa_i, sub_{xy}, \lambda_C^{++}(sub_{xy}), \lambda_C^{+-}(sub_{xy}), \lambda_C^{-+}(sub_{xy}), \lambda_C^{--}(sub_{xy})]$
$\models_G$ **possible**$([d]$ is about $[sub_x]$ and $[sub_y])$
holds if and only if $\lambda^{\wedge}_{\mathbf{minPos}} \leq \lambda_C^{++}(sub) < \lambda^{\wedge}_{\mathbf{maxPos}}$.

$conholon[\kappa_i, sub_{xy}, \lambda_C^{++}(sub_{xy}), \lambda_C^{+-}(sub_{xy}), \lambda_C^{-+}(sub_{xy}), \lambda_C^{--}(sub_{xy})]$
$\models_G$ **believing**$([d]$ is about $[sub_x]$ and $[sub_y])$
holds if and only if $\lambda^{\wedge}_{\mathbf{minBel}} \leq \lambda_C^{++}(sub) < \lambda^{\wedge}_{\mathbf{maxBel}}$.

$conholon[\kappa_i, sub_{xy}, \lambda_C^{++}(sub_{xy}), \lambda_C^{+-}(sub_{xy}), \lambda_C^{-+}(sub_{xy}), \lambda_C^{--}(sub_{xy})]$
$\models_G$ **knowing**$([d]$ is about $[sub_x]$ and $[sub_y])$
holds if and only if $\lambda_C^{++}(sub_{xy}) = \lambda^{\wedge}_{\mathbf{Know}} \mathbf{= 1}$.

For purely editorial reasons, we do not deal with the complementary conjunctive alerts, i.e., alerts on new documents *being about* $[sub_x$ *and not* $sub_y]$, $[$*not* sub_x *and* $sub_y]$, $[$*not* sub_x *and not* $sub_y]$. It is quite obvious that they have to be verified in a similar way, but against values of $\lambda_C^{+-}(sub_{xy})$, $\lambda_C^{-+}(sub_{xy})$, and $\lambda_C^{--}(sub_{xy})$, respectively.

3.3 A Brief Note on Cognitive Semantics

The novelty of our approach to the generation of quasi-natural language alerts falls outside of previous linguistic models. Namely, it is an original proposal consistent with cognitive linguistics (Evans and Green, 2006) and interactive linguistics (Wlodarczyk, 2013) paradigms. Both of them refer our work to the concept of cognitive semantics (Talmy, 2000), which describes the way a particular natural language sentence embraces the pre-linguistic knowledge corpora accessible to minds of a communicative agent. Obviously, in our *R&D* context the communicating subjects are IF systems.

Cognitive semantics is always characterised by high specificity, because in each case it reflects pragmatics and meaning of a very narrow class of linguistic phenomena. In our model this specificity is apparently visible in internally related and complex structure of mental language holons. A proposal of how to realize the cognitive semantics of alerts in our IF system should be treated as the most original contribution of the model.

4 COMPUTATIONAL EXAMPLE

In this section we introduce a basic example that illustrates the entire process of generating linguistic alerts in IF systems. For the sake of simplicity let us assume an elementary information systems comprised of a document repository consisted of 10 processed documents $D = \{d_1, d_2, ..., d_{10}\}$ and 3 new documents $D^{new} = \{d_{11}, d_{12}, d_{13}\}$ that are evaluated based on a set of 4 conditional attributes $W = \{w_1, w_2, w_3, w_4\}$. Further, let us assume that user's information needs are limited to two subjects $S = \{sub_1, sub_2\}$. Consequently, the set of all attributes available in the system is defined as $A = \{w_1, w_2, w_3, w_4, sub_1, sub_2\}$. Furthermore, let the domains of the introduced attributes be given as follows, $W_1 = W_2 = W_3 = W_4 = \{A, B, C\}$.

Documents stored in the document repository are processed. In particular, each document is analysed by a set of indexing mechanisms (or other processing mechanisms) that are able, based on the document content and structure, to assign values for each of the conditional attributes. Further, information about each document's subject is determined and stored. As such the information function of the repository is determined, i.e., attribute–value mapping, as given in Table.1.

Focusing on three simple classes κ_1, κ_2, and κ_3, given as $\kappa_1 = \{(w_1, B), (w_2, A), (w_3, A), (w_4, A)\}$, $\kappa_2 = \{(w_1, C), (w_2, C), (w_3, A), (w_4, B)\}$, and $\kappa_3 = \{(w_1, B), (w_2, B), (w_3, C), (w_4, A)\}$, we can determine three non-empty clusters of documents $K_1 = \{d_1, d_2, d_3, d_6\}$, $K_2 = \{d_4, d_5, d_7, d_{10}\}$, $K_3 = \{d_8, d_9\}$ and their extensions $EXT(K_1) = \{d_{11}\}$, $EXT(K_2) = \{d_{12}\}$, $EXT(K_3) = \emptyset$. It must to be mentioned that one of the newly received documents, namely d_{13}, does not belong to any of these sets. This fact will be commented in the final remarks section.

Resulting summarization of data is represented by the following set of holons HOLONS = SIMHOLONS ∪ CONHOLONS:

$$\begin{aligned} SIMHOLONS = \{ & \\ & simholon[\kappa_1, sub_1, 0.25, 0.75], \\ & simholon[\kappa_1, sub_2, 1.00, 0.00], \\ & simholon[\kappa_2, sub_1, 1.00, 0.00], \\ & simholon[\kappa_2, sub_2, 0.25, 0.75], \\ & simholon[\kappa_3, sub_1, 0.50, 0.50], \\ & simholon[\kappa_3, sub_2, 0.00, 1.00]\}. \end{aligned} \tag{10}$$

$$\begin{aligned} & CONHOLONS = \{ \\ & conholon[\kappa_1, sub_1 \wedge sub_2, 0.25, 0.50, 0.25, 0.00], \\ & conholon[\kappa_2, sub_1 \wedge sub_2, 0.25, 0.75, 0.00, 0.00], \\ & conholon[\kappa_3, sub_1 \wedge sub_2, 0.00, 0.50, 0.00, 0.50]\}. \end{aligned} \tag{11}$$

Having the relative grounding strength computed and stored in each holon, we can now determine all proto-forms, for the new arrivals from non-empty extensions $EXT(K_1)$ and $EXT(K_2)$.

To give an example, simple subjects will be considered. Let modality thresholds be set up to following values $\lambda^{KS}_{Know} = \lambda^{KS}_{maxBel} = 1$, $\lambda^{KS}_{minBel} = \lambda^{KS}_{maxPos} = 0.60$, and $\lambda^{KS}_{minPos} = 0.20$. These values are not accidental. Namely, they have been chosen taking into account theorems from the theory of grounding simple modalities (Katarzyniak, 2005). It follows that the threshold values should preserve consistency of sets of grounded proto-forms with common sense interpretation. Below we provide examples of well-grounded grounded proto-forms:

Table 1: Processed repository of documents.

	w_1	w_2	w_3	w_4	s_1	s_2
d_1	B	A	A	A	1	1
d_2	B	A	A	A	0	1
d_3	B	A	A	A	0	1
d_4	C	C	A	B	1	0
d_5	C	C	A	B	1	0
d_6	B	A	A	A	0	1
d_7	C	C	A	B	1	0
d_8	B	B	C	A	0	0
d_9	B	B	C	A	1	0
d_{10}	C	C	A	B	1	1
d_{11}	B	A	A	A	ε	ε
d_{12}	C	C	A	B	ε	ε
d_{13}	B	B	C	B	ε	ε

- **possible**([d_{11}] is about [sub_1]) AND **possible**([d_{11}] is **not** about [sub_1])
- **believing**([d_{11}] is about [sub_2]), BUT STILL **possible**([d_{11}] is **not** about [sub_2])
- **knowing**([d_{12}] is about [sub_2])

It is worth of mentioning that these proto-forms are logically consistent, which is ensured by the proper choice of modality thresholds. A possible natural language alert founded on the established proto-forms is:
IF SYSTEM ALERT: There is a new [document: doc_{12}] available. I believe it is about [subject: sub_2]. You may be interested in reading it!.

5 FINAL REMARKS

The theoretical foundation for designing and implementing interactive IF systems is proposed in this paper. The desirable common sense consistency of quasi-natural language alerts is ensured by the application of a theory of epistemic modality grounding, introduced elsewhere. The proposal substantially differs from previous models of similar alerts generation.

The proposed model of linguistic alerts choice and production is supported by a simple computational methodology and a naive model of uncertain classification rules. Alternative and more sophisticated approaches are possible (and required) e.g. for the way sets K_i and $EXT(K_i)$ are determined. Obviously,

complete final implementation need to cover the missing case of document d_{13}, either.
The introduced model supports effective design and implementation of modern interactive and mobile tools for alerting end users about newly received objects of potential interests, in both written and vocalized modal natural languages.

ACKNOWLEDGEMENTS

This work was realized under research cooperation between Wrocław University of Technology (Faculty of Computer Science and Management - Internal Grant No. S50198 K0803) and Hradec Králové University (Center for Basic and Applied Research, Faculty of Informatics and Management SP-FIM-2016 - Smart Solutions for Ubiquitous Computing Environments).

REFERENCES

Brown, P. J. and Jones, G. J. F. (2001). Context-aware retrieval: Exploring a new environment for information retrieval and information filtering. *Personal and Ubiquitous Computing*, 5(4):253–263.

Evans, V. and Green, M. (2006). *Cognitive linguistics: An introduction*. Edinburgh University Press.

Hanani, U., Shapira, B., and Shoval, P. (2001). Information filtering: Overview of issues, research and systems. *User Modelling and User-Adapted Interaction*, 11(3):203–259.

Herrera-Viedma, E., Herrera, F., Martínez, L., Herrera, J. C., and López, A. G. (2004). Incorporating filtering techniques in a fuzzy linguistic multi-agent model for information gathering on the web. *Fuzzy Sets and Systems*, 148(1):61–83.

Johnson-Laird, P. N. (1985). *Towards a Cognitive Science of Language, Inference, and Consciousness*. Cambridge University Press, Cambridge.

Katarzyniak, R. (2005). On some properties of grounding simple modalities. *Systems Science*, 31(3):59–86.

Katarzyniak, R. (2006a). On some properties of grounding nonuniform sets of modal conjunctions. *Int. Journal of Applied Mathematics and Computer Science*, 16(3):399–412.

Katarzyniak, R. (2006b). On some properties of grounding uniform sets of modal conjunctions. *Journal of Intelligent and Fuzzy Systems*, 17(3):209–218.

Lorkiewicz, W., Popek, G., Katarzyniak, R., and Kowalczyk, R. (2011). Aligning Simple Modalities in Multi-agent System. In *Proc. ICCCI 2011*, volume 6923, pages 70–79. LNAI.

Nuyts, J. (2001). *Epistemic Modality, Language, and Conceptualization: A Cognitive-pragmatic Perspective*.

Pawlak, Z. and Skowron, A. (2007). Rudiments of rough sets. *Information Sciences*, 177(1):3–27.

Shapira, B., Shoval, P., and Hanani, U. (1999). Experimentation with an information filtering system that combines cognitive and sociological filtering integrated with user stereotypes. *Decision Support Systems*, 27:5–24.

Stachowiak, F. J. (2013). Tracing the role of memory and attention for the meta-informative validation of utterances. In Wodarczyk, A. and Wodarczyk, H., editors, *Meta-informative Centering in Utterances: Between Semantics and Pragmatics*, pages 121–142. John Benjamins Publishing Co., Amsterdam.

Talmy, L. (2000). *Toward a Cognitive Semantics*. MIT Press, Cambridge, MA.

Wlodarczyk, A. (2013). Grounding of the meta-informative status of utterances. In Wlodarczyk, A. W. and H., editors, *Meta-informative Centering in Utterances: Between Semantics and Pragmatics*, pages 41–58. John Benjamins Publishing Co., Amsterdam.

Xu, M., Ong, V., Duan, Y., and Mathews, B. (2011). Intelligent agent systems for executive information scanning, filtering and interpretation: Perceptions and challenges. *Information Processing and Management*, 47(2):186–201.