Embedded fuzzontological model for document interpretation in attribute-value-pair annotation in economic intelligent systems

Olufade Onifade, Olusoji Okunoye, Amos David

EMBEDDED FUZZONTOLOGICAL MODEL FOR DOCUMENT INTERPRETATION IN ATTRIBUTE-VALUE-PAIR ANNOTATION IN ECONOMIC INTELLIGENT SYSTEMS

Olufade, F. W. ONIFADE
University of Ibadan, Dept. of Computer Science, Ibadan, Oyo State, Nigeria
LORIA-Campus Scientifique, BP 239 – 54506, Vandoeuvre, Les Nancy, France
Olusoji, B. OKUNOYE
University of Lagos, Akoka, Lagos, Nigeria
LORIA-Campus Scientifique, BP 239 – 54506, Vandoeuvre, Les Nancy, France
Amos DAVID
LORIA-Campus Scientifique, BP 239 – 54506, Vandoeuvre, Les Nancy, France

Abstract: Two popular views about the concept of information are the "information as a process" and "information as a product". While there has been much research about the two paradigms, one recurrent question has been where is the place of interpretation in information utilization? A growing trend is the concept of annotation which encourages individual interpretation on a subject of interest and keeps such information for future use. The problem of interpretation is not new, as it was clearly pointed out in the infological equation. With several attempt at resolving this imbroglio in place, we are of the opinion that misinterpretation results from differences in individual knowledge and cognitive ability amongst other factors. Consequently, we developed an attribute-value-pair (AVP) document representation metadata to interface directly with the ontology domain, and extend interpretation via fuzzy inference system.

Keywords: Fuzzontology, Attribute-Value-Pair Annotation, Information need, Document interpretation, Economic intelligence

1. Introduction

Wang et al. (1998) was amongst the earlier proponent of the concept in which information as a product was greatly elaborated. Information product (IP) is defined as a collection of data element instances meeting the specified requirements of a data consumer, these requirement are usually employed for business decision making, legal reporting or government reporting (Lee et al. 2002). In treating information as a product the following must be taking into cognizance: understand the consumer's information need, manage information as a product of well-defined production process, manage information as a product with life cycle, and appoint an information product manager to manage the information product. A similar notion to the above is to view information as a tangible resource or a product which is manufactured (Chaffey & Wood, 2005).

A contrasting view to the notion of information as a product is the consideration of information as a process. It emphasizes human role in the creation of information. Popular proponents are Mutch (1996), Davenport (1997) and the work of Davenport & Prusak (1998).

However, the above two paradigm did not give much attention to the result of users interpretation in the usage of information. In demonstrating the workability of the 'infological equation', Langefors (1966, 1995) was of the opinion that those who are to interpret data in order to inform themselves must be viewed as part of the system. Using the equation I = i (D, S, t), where 'I' is the information (knowledge) produced by a person from the data 'D' alongside with pre-knowledge 'S' through an interpretation process 'i' in interval 't'. Bednar & Welch, (2007) reflecting on the opinion of Langefors made by Schutz, (1967) reiterated the impossible nature of communicating "meaning" between people. This was sequel to the fact that simply transmitting data will not lead to communication of shared understanding knowing fully well that 'i' and 'S' cannot be assumed to be common. To this end, communication can only be seen to approach success most closely where individuals interpreting the same data belong to a group with possible vested professional interest. This is sequel to the fact that every act of interpretation does not necessarily invoke the entire 'S' attribute to every individual, thereby creating room for some aspect of shared experience leading to similarities in the 'i' among group members (Bednar & Welch, 2007).

It is not a common place for a piece of data to generate similar 'factual' meaning when interpreted by different individuals. However, derivable inferences would be likely different more widely in 'meaning' of the data for different individual based on his/her associations, and/or possible consequences depending on the uniqueness of 'S'. Communication and intention is context-dependent. Interpretation of context continually evolves with time thus having great influence on sense-making and communication (by Wittgenstein, (1963) and quoted in Bednar & Welch, (2007)).

With annotation making waves as a means of representing and improving users ability to express their observation and contextualise it, the issue of interpretation when several individual are concerned has not been properly addressed. The subjective nature of information and its usage for information need thus require a more flexible system that can harmonize the common vocabularies in a domain and employs adaptive means of resolving the ambiguity in natural languages.
In the rest of this paper, we discuss economic intelligence and present the intelligence cycle in section two. We present our notion of AVP annotation in

section three within the realm of economic intelligence. In section four, we present the modified fuzzontological model to accommodate the AVP metadata. Section five concludes this research and gives directions for our future work.

2. Economic Intelligence (EI) & the Intelligence Cycle

EI can be described as a set of *coordinated actions* of *search*, *processing* and *distribution for exploitation* of *useful information* for *economic actors* (Martre, 1994). A proper juxtaposition of the definition depicts a set of mappings: the whole process can be described as a set of "*coordinated actions of search and processing"* – these activities implies that they are not limited to search from the information delivery systems alone but also encompass the brain tasking activities for the acquisition of related experiences and other cognitive factors (Wang et al. 2004). Similarly, the "*distribution for exploitation of useful information"* although literarily simple, but involved a herculean task of rigorously defining and redefining of the earlier stages to guarantee a proper understanding for the rationale for the decision making.

The three principal concepts drawn out of the various definition of EI are the user, process and information which are inherent in all forms of human interactions necessitating decisions. It therefore implies that the understanding of their individual operation cum interaction will go a long way to enhance the overall objective as depicted in the definitions earlier presented.

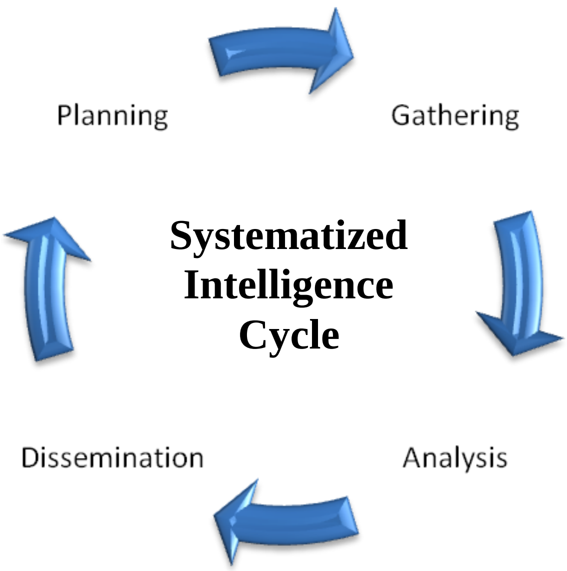

Figure 1: The Intelligence Cycle

Appropriate information on a timely and credible manner facilitates intelligence in the delivery of strategic decision. As shown in figure 1, the intelligence cycle usually starts with planning or what we normally refer to a definition of decision problem by the decision maker. A properly defined and scrutinized decision problem leads to definition of users' information need for the particular problem. It should be stressed that, neither of these stages is independent of human actors and thus the subjective nature which forms the basis of this research. Once the information need has been defined, the process of gathering commenced. In EI system, the formulation of information need is relative to the understanding garner from the deliberation between the actors involved.

Once a sizeable or required volume of information has been acquired, the process of analysis is the next stage. Herein, factors like validation of information and the mode of presentation for final use are of importance. Information dissemination can either be directly to the public or utilized for the earlier defined decision problem. In all, this forms a cycle which evolves as the problem is considered, information gathered and analyzed.

In an attempt to assist both the information watcher (information specialist) and the decision maker access to time, usable information, and the possibility of reusing of existing information, we propose the AVP annotation model to capture the divergent opinions of economic intelligence actors. Next section summarises our proposition on AVP annotation.

3. Attribute Value Pair (AVP) Annotation

Attribute-Value Pair (AVP) can be described as a form of data representation in computer systems. Data element can be represented by a set of attributes and values. In this context, the term attribute means an abstraction, property, or a characteristic of an entity or an object. For example, an object car has colour attribute. The attribute value could be red. Ability of a user to express his observation and/or contextualize document object of interest as attribute-value pair annotation could improve significantly the effectiveness of such valued-added information. It will provide a good basis for data restructuring, data mining, robust exploitation, knowledge elicitation among others. The above has been tested in our earlier work in which we employed this approach (annotation representation as AVP) to assist in materializing knowledge in knowledge capitalization (Onifade, et al., 2009).

The motive to represent annotation as attribute-value pair is born from the fact that the needs of users are evolving. We cannot pre-conceive all possible needs in the design of information system. The system design is targeted towards achieving a particular goal. Users are constraint to assign values to only available attributes. The granularity level of attribute value a user need may not be available. Where a user specifies an attribute that does not exist, there would be a need for structural changes to the schema of the Information System in order to accommodate such new attributes. Such changes might be tedious, time consuming and costly to effect. Annotation as attribute-value pair is a way of solving the problem of structural changes without modifying the underlying schema. The depositor of formal information (information published on the web) for example, determines the information contents based on his initiative and perception of the domain. The users, on the other hand, use the available information to meet their needs. There is a possibility of a gap to exist between the goal of the user and the goal of the depositor (or author) as a result of lack of correspondence between the information contents and the user needs.

Two *forms* of annotations can be inferred: annotations made on the entire document (referred to as A_{T1}) and annotations made on the objects of the document (referred to as A_{T2}). The objects of document such as terms, phrase, sentences, etc will be called *document-objects (do)*. In this phase, watcher might give the meaning or definition of each term identified; the synonyms of the term and the source in form of annotations. At this point, it is important to differentiate between the process of adding annotation to a document and the actual annotation added. While the former is regarded as annotation-process ($A_{procees}$), the latter is annotation representation as AVP. This is denoted as A_{avp}. A_{T1} and A_{T2} are form of A_{avp}. The set A_{avp} of annotation is defined as

$$A_{avp} \subseteq id \times def \times attr \times val \times timestamp \qquad (1)$$

Where *id=identifier, def = definition, attr = attribute name, val = annotation value*

Bouaka in [12] proposes DMP model for structuring decision problem. She translated decisional problem in terms of stakes. A stake is a set of object-environmental object detected and proposed by decision maker, signal – the meaning the decision maker gives to the detected object, and hypothesis – the possible result or outcomes associated with each signal. What constitutes the object may be inferred from the document. However, the signal(s) and the hypotheses are derivables based on the overall objectives of solving the decision problem. Again, annotations can be used to add to the model, what constitute the signal(s) and hypotheses.

In information collection, watcher searches through available information sources in order to identify and collect possible information that could relevant to solving the decision problem. There may be need to structure or restructure the collected information, reference the sources, add attributes and/or values. These tasks may be performed with the aid of annotation. Annotation can also be used to determine the relevance of retrieved information sources relative to decisional problem being solved. It is important to note here that time and the context of the problem being solved would greatly influence the relevance of a document. In the extraction of possible indicators for decision making from the retrieved information sources, annotation could be used to add user's interpretation of the content of documents.

Economic Intelligence (EI) actors may share their perceptions of the decision problem or of any information through annotation in a collaborative environment. The conclusion arrived at about the object of deliberation (decision problem) could be added to the initial information source as annotation. The use of annotation as a platform for economic actors to share one another's perception could be done synchronously or asynchronously. With synchronous annotation, users can communicate with one another in real time. It however, requires all participating actors to be online. However, with asynchronous annotation, the actors need not communicate in real time. Electronic mail or any other form could be used in the communicative acts. It however, requires a means of notifying the involving actors about the pending message as time may be of essence. Therefore, the design and development of annotation model for economic actors (users) that can provide annotation capabilities for adding values to information sources are important as well as very necessary. The overall structure is illustrated in figure 2 below.

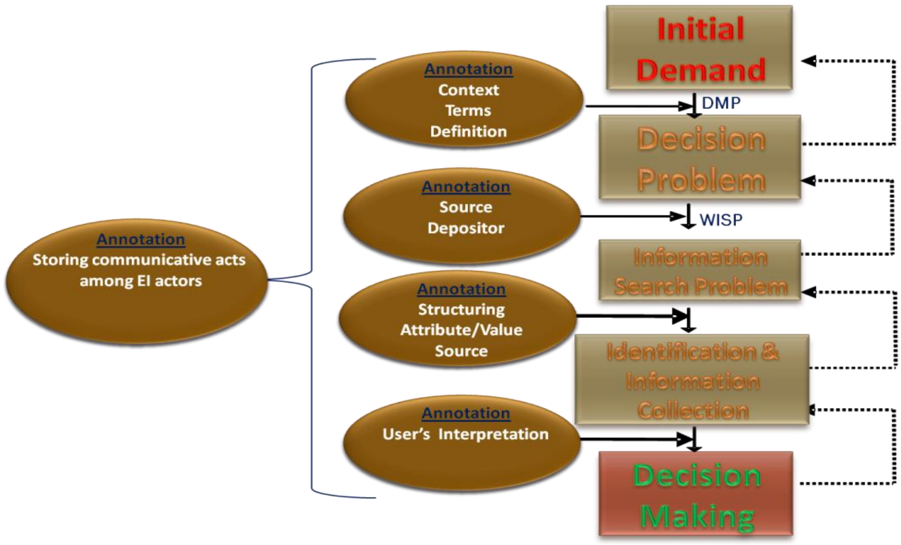

Figure 2: Relevance of Annotation in Economic Intelligence

A decision problem often starts with an initial demand from the decision maker based on some perceived signals. The initial demand, produced as document D is a materialized knowledge of the decision maker on the problem to be solved. In the transformation of this initial demand into a decision problem, the actors often need to understand some basic things about the problem such as the context or environment of the problem if they are present in the initial demand. Annotation could be used by watcher to add his interpretation to the document (initial demand). Decision maker may validate watcher's interpretation through annotation, however robust means of achieving this is imperative in a multi-actor interpretation scenario.

4. Fuzzontology

Information mining is the non-trivial process of identifying valid, novel, potentially useful and understandable patterns in heterogeneous information sources. Considered as an offshoot of knowledge discovery in databases (KDD) or data mining, it is saddled with the exploitation of information for decisional purposes. However, it's without gainsaying to assert that language is human most effective tool to structure his experience and also model his environment. Langefors, (1966) in his equation stressed the importance of interpretation which is peculiar to human beings. It is therefore important to model linguistic terms similar to the work of Zadeh in "Computing with words" (Zadeh, 1996). Fuzzy inference systems have been applied successfully in many fields which include automatic control, data classification, decision analysis, expert systems and computer vision to mention a few.

Complexity of problem definition derives from the differences in interpretation of each key actor in a given situation usually results into different level of comprehension of the event and to the explanations of the influences between events. In the formal part, it is important to recognize that personal values play a part in interpretation, and the latter part opined that individuals brings to bear different experiences and wisdom that has created different belief systems (Shankaranarayanan, et al., 2003). An interpretation is in itself subjective, i.e. related to a subject, prevailing circumstances, a mind, ego, or agent of whatever sort that sustains or assumes the form of thought or consciousness.

Wand & Wang (1996) are one of the key proponent of ontological dimension to data quality employed a formal model for an information system by considering mapping function from the real world to an information system. Bearing in mind the need to reduce the risk of non-quality data during information search operation towards the delivery of strategic decision making in economic intelligent systems, we proposed ontological framework for knowledge reconciliation to facilitate proper understanding between the decision maker (DM) and the information specialist (Watcher) the duo of which are identified and established actors in economic intelligent process.

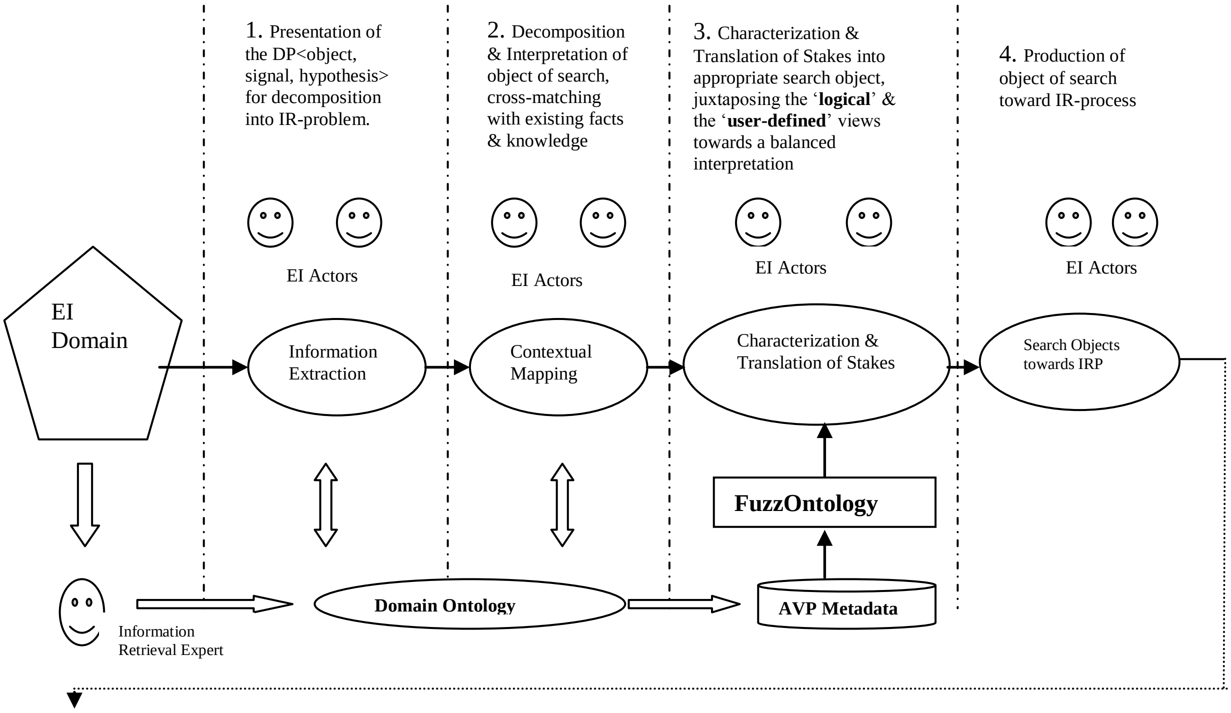

Figure 3: Embedded FuzzOntological Model with AVP Annotation

We define Knowledge reconciliation as the attempt to map the desire for precision in naturally expressed languages employed in describing, sharing of knowledge, interpreting and communication of "a need" to another person/object in an acceptable degree of accuracy devoid of misinterpretation, disinformation, biases and unnecessary personal preferences to mention a few, which was referred to as risk factors (RFs) in (Onifade, et al., 2008). This becomes imperative because real world is laden with concepts which do not have a sharp boundary e.g. 'fine', 'useful', 'more important than', 'old' e.t.c. We therefore employed this background to develop an ontological framework aimed at reducing the risk in the knowledge reconciliation's operations. A detail discussion on the KNOWREM framework can be found in *(Onifade, et al., 2008).*

With this background, we present the main contribution of this work in the next section where we employed the same framework in figure 3 to improve the judgment of the actors involved via the inclusion of fuzzy inference system to facilitate robust decision making in "FuzzOntology".

The importance and applicability of our model in figure 3 named fuzzontology was articulated in Onifade, et al., (2010). We therein established its potential facilitate problem evaluation amidst multiple actors. This ability is not comparable with the provision of tools like SWOT and SACH but also provides a very flexible manner of representing and constructing the analysis stage.

Ontological application in risk simulation brings about the conceptual requirement of “reusability”. This is sequel to the fact that models are specified by humans to embody domain knowledge, characterized by ambiguity, thus ontology provides a profound solution in bridging the semantic-gap between the knowledge-space and the simulation model [Alan, et al., 2007]. In figure 3, the domain ontology therefore, the domain ontology establishes the common vocabularies among the economic intelligent actors, while the AVP metadata represents the added annotation from this actors. Usually, the problem arises when more than one inference can be legitimately derived from a word or statement more so in the absence of the originator.

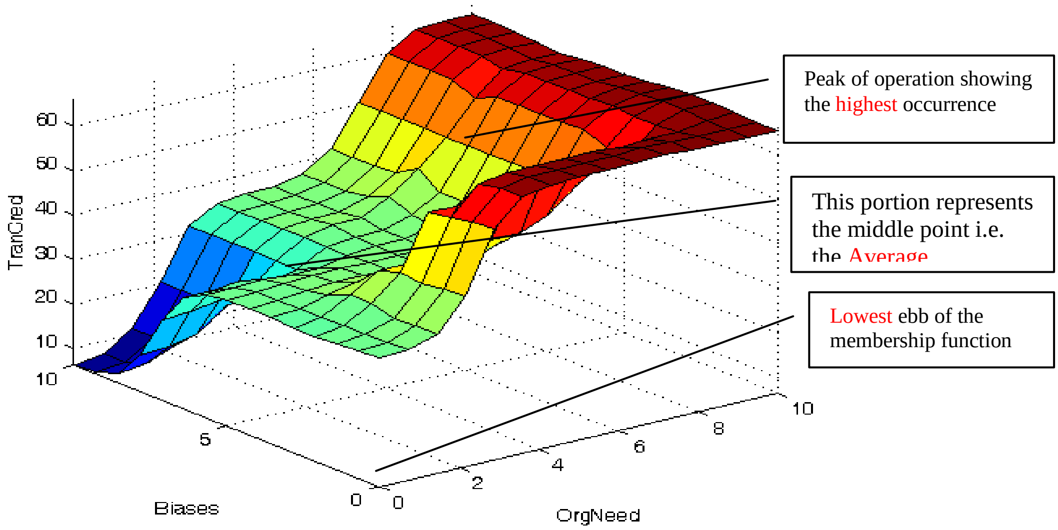

Figure 4: Fuzzified result for EnviFac and OrgNeed

Most decision process and result evaluation fails to include the effect of biases and personal preferences. While many of these procedures recognized the importance of these intangible factors in effective delivery of decision, capturing and utilizing them becomes a serious challenge. Experiments showed that a set of judges evaluating student moral and academic performance usually attempt to find something good to say about their student once they are familiar or know the student – bias, but to others, they tend to be more focused and accurate in their judgment.

We developed a fuzzy inference system to interpret words with uncertain and ambiguous meaning from the domain ontology and AVP annotation metadata. Several factors were considered from the decision making deliberation point. We look at factors usually analyzed with decision analysis tools, and determine their effects on the overall functionality of the system. An interesting part of this contribution is that we can also include intangible factors which are hitherto difficult to include and are therefore neglected in rational decision making. Figure 4 is the fuzzified result of articulating the interpretation of the actors on the organization's need as viewed by each of them. We implicitly add an intangible factor which is not easy to capture like bias and study the resultant effect on the overall implication of interpretation provided by the actors.

We include root of biases and personal preferences as being integral part of human decision making. This includes other factors like favouritism whose effect are not only difficult to measure but also to capture. Bias is naturally taken as being negative, and towing this line of reasoning we can see a sharp turning in figure 4 as opposed to the other two figures. The prominent rate of bias displayed forces the blue colour to the high side of the chart. Clearly captured is the fact that, not minding the level of understanding of the organizational need presence of bias really distorts the decision making process. The prevalent rate of bias could not allow the knowledge about the organization displayed to have any significant effect. It is again interesting to see that if it is almost possible to remove bias, the *OrgNeed* stay put at half the chart shown in green colour. The gradual increase in the value of the green (moderate/average) factors is observable with its inclusion around 7 on the bias axis.

5. Conclusion

Interpretation is the driving factors in the delivery of any logical decision. While the AVP annotation enhances knowledge reutilization, we notice the problem of interpretation in the reuse of such knowledge. Following closely the infological equation; we observed that no two individuals can interpret the same information the same way. Most of available information is subjective if not biased. How then can there be proper resolution in the subjective judgment if it is not treated fuzzily. Fuzzontology employs similar factors and developed a membership function for the environmental factors which might be difficult to be expressed precisely. Apart from this tangible factor, we also model and simulate the inclusion of intangible factor like biases and personal preferences which cannot be captured in existing tools. In the future, we hope to establish a pattern of usage by user via the interpretation derived from Fuzzontology.

Reference

Allan,G. M N.D. Allan, V. Kadirkamanathan & P.J. Fleming "Risk Mining For Strategic Decision Making": K.M Wegrzyn-Wolska and P.S. Szazcepaniak (Eds): Adv. In Intel. Web, ASC 43, pp 21 – 28, 2007, Springer-Verlag Berlin Heidelberg.

Bednar, P. M. & Welch, C. (2008): "Bias, Misinformation and the Paradox of Neutrality" In Informing Science: the International Journal of an Emerging Transdiscipline Volume 11, 2008, pp 85 - 106

Chaffey, D. & Wood, S. (2005). *Business information management: Improving performance using information systems.* Harlow, Prentice Hall, Pearson Education Limited.

Davenport, T. H. (1997). *Information ecology*. Oxford: Oxford University Press.

Davenport, T. H., & Prusak, L., (1998) "Working Knowledge: How organizations manage what they know". Boston – Harvard Business School Press

Lee, Y. W., Strong, D. M., Kahn, B. K. & Wang, R. Y. (2002): AIMQ: A Methodology for Information quality assessment. Elsevier Information and Management 40 (2002) 133 – 146.

Onifade O.F.W. , Thiery O., Osifisan., A.O. & Duffing G. (2008b): " Ontological Framework for Minimizing the Risk of Non Quality Data during Knowledge Reconciliation in Economic Intelligence", in proc of the 13th Int'l Conference in Information Quality (ICIQ'08), MIT, Boston, pp 296 – 309.

Onifade O.F.W. , Thiery O., Osifisan., A.O. & Duffing G. (2010): "FUZZONTOLOGY: Resolving Information Mining Ambiguity in Economic Intelligent Process". S.K. Prasad et al. (Eds.): ICISTM 2010, CCIS 54, pp. 232–243, 2010. Springer-Verlag Berlin Heidelberg 2010

Onifade,O.F.W., Oladejo, B. F. & Okunoye, O. B. (2009): Risk Indicators as a Source of Knowledge in Economic Intelligence (EI). Symposium d'analyse et de réflexion sur les: Models, methods, Ingéniere de l'intelligence Compétitive. Beaulieu Sur Mer – 25, 26 Nov 2009

Redman, T.C. (1998): "The Impact of poor Data Quality on the Typical Enterprise". Communications of the ACM, Volume 41, Issue 2 (Feb. 1998), Pages: 79-82

Shankaranarayanan G., Ziad M., & Wang R. Y. (2003), "Managing Data Quality in Dynamic Decision Making Environments: An Information Product Approach", Journal of Database Management (14:4), Oct – Dec 2003, Pp. 14 – 32.

Wand, Y. & Wang, R.Y. "Anchoring Data Quality Dimension in Ontological Foundations". Communications of the ACM, November 1996, 39(11), 86-95.

Wand, Y. & Weber, R.: "On the Ontological Expressiveness of Information Systems Analysis and Design Grammars". Journal of Information Systems (1993) 3, Pp. 217 – 237.

Wang, H., Johnson, T. R. & Zhang, J. (1998): "UEcho: A Model of Uncertainty Management in Human Abductive Reasoning". In proc. Of 20th Annual Meeting of the Cognitive Science Society, 1998, Hillsdale, NJ

Wang, Y., Wang, Y., Patel, S. & Patel, D. (2004) "A Layered Reference Model of the Brain", IEEE Transactions on Systems, Man and Cybernetics, vol. 34

Zadeh L. A., The concept of a linguistic variable and its application to approximate reasoning, parts I, II, Information Science, 8, pp. 199–249, pp. 301–357, 1975.

Zadeh, L. A.: Fuzzy Logic= Computing with words (CW) "IEEE transactions on fuzzy systems, vol 4, no 2, May 1996, pp 103-111.