From Ontologies to Trust through Entropy

Anusch Daemi and Jacques Calmet
IAKS Calmet
Universität Karlsruhe (TH)
Am Fasanengarten 5, 76131 Karlsruhe
Germany
Email: {daemi, calmet}@ira.uka.de

***Abstract*— We propose a methodology to identify structured knowledge (ontologies) that is based upon the concept of entropy and its associated distance measures, i.e. the Kullback-Leibler distance or mutual information. At the heart of such a methodology is the probability distribution of states, where the states define the knowledge for a given application field. The first approach presented in this paper either identifies a new structure or extends an existing structure on a taxonomy. In contrast to Resnik's similarity measure, used in computational linguistics, we do not consider probability distributions of large corpora of text or any documents at all. In a first step, we impose a probability distribution on the elements of an ontology, where the distribution depends on the constraints set for the intended meaning of the ontology.**

I. Introduction

On one side, there exist some concepts that ought to play a key part in the validation and trust of E-transactions. They range from degrees of belief in practical reasoning to the very nature of complexity as pointed out by Gell-Mann [1] when he defines crypticity as the way to extract a theory from a set of data, covering also numerical statistics or classification as used by insurance companies. Each of these approaches is at the center of a well-defined research field mostly disconnected from the others. On the other side, we deal daily with the concept of ontology [2] as it is used for instance for the Semantic Web [3]. Usually, the concept of ontology is defined as a kind of taxonomy or even more crudely as a list of words describing domain specific knowledge. A proper definition of ontology though implies that the knowledge is structured, which is also consistent with the usual concept first introduced by philosophers. This work is an attempt to set ontologies in the framework of validation. We propose a methodology to identify structured knowledge (ontologies) and to assess what can be seen as a trust (or classification) feature. It is based upon the concept of entropy and its associated distance measures, i.e. the Kullback-Leibler distance. At the heart of such a methodology is the probability distribution of states, where the states define the knowledge for a given application field. In a previous paper [4] an extension of the concept of distance to ontology, arising from entropy, was already described but without proposing a general methodology to identify ontologies.

The proposed method reminds of many others in various fields ranging from numerical statistics and classification to non-monotonic reasoning and degrees of belief. In the latter there are several approaches (see i.e. [5], [6], [7], [8]), one of them using the notion of entropy. This requires for example the computation of entropy according to the principle of maximum entropy. One of the contributions of this paper is to compute only a distance function to achieve similar results, but more efficiently.

It should be mentioned that the methodology shown above is completely different from those used in computational linguistic to define some kind of distance. There one has to make use of large corpora of text, for example the Brown Corpus of American English [9], to define a distribution on the concepts in a taxonomy. With these probability assignments a similarity measure is calculated, like the well know Resnik measure [10], or some advanced measures like *tf/idf* (for example [11], that are more or less based on Resnik's notion of information content).

The paper is structured as follows. In section II we will give a brief overview over some distance measures in computational linguistic and a succinct history of the concept of entropy. Section III details the way how we can identify a structure in the classical sense when we have data as information sources, for example simple numbers from a measurement. This model is extended to identify a structure on arbitrary knowledge and hence allow for an identification of ontologies. This is illustrated with a very simple example on a taxonomy of musical styles. Section IV shows some application examples for the newly introduced distance measure and the imposed structure. In section V we outline the link between trust and ontologies. The paper concludes with section VI, where we give an outlook for ongoing research.

II. Distance and Entropy

The following section presents a short introduction to distance measures in computational linguistic, especially Resniks measure as we need it for comparison purposes. Then, an overview of entropy is given, as it is essential for our model of identifying ontologies.

A. Distance in computational linguistic

For measuring the distance of special discrete structured objects like words, documents and similar corpora of text, computational linguistic has introduced, among others, the distance measures *positioning* and *frequency*.

The positioning distance measure is frequently used in Vector Space Models (VSM) where documents and queries are represented in a high dimensional vector space. The basis vectors are composed of index terms, which are words relevant to the problem. The location of the document vector in the VSM is determined by weights assigned to the index terms for each document [12]. These weights can be calculated in different ways. Usually they are either term frequencies (*tf*) or inverted document frequencies (*idf*). The distance between a document and the query is consequently measured using the cosines between the two vectors (also known as normalized correlation coefficient [13]). The selection of basis vectors for the VSM is critical for successful recognition of similarities between documents, because they can severely warp the distance measure if poorly chosen.

The frequency approach is based on the frequency of terms in a document, term frequency, which is used in the information retrieval measure *tfidf*. Term frequency/inverted document frequency weights the term frequency of a term in a document by a logarithmic factor that discounts the importance of the term concerning all relevant documents. So, terms appearing too rarely or frequently have a lower weight than those holding the balance [11].

A distance measure originating from term frequency and *tfidf* can be calculated by introducing some form of entropy as first presented by Resnik in [10]. This method assigns to instances of concepts c a probability $p(c)$ for encountering them in a document. The information content of such a concept c is subsequently defined as $-\log p(c)$. The distance measure is based on the assumption that the more information two concepts have in common, the more similar they are. The information shared by two concepts is indicated by the information content of the concepts that subsume them in a given taxonomy. In practice, one needs to measure similarity between terms rather than between concepts, so calculations have to be done over sets of concepts representing meanings of the terms.

B. Overview of entropy

The concept of entropy originates in physics through the second law of thermodynamics [14]. Another important area where entropy plays a central role is statistical mechanics, which is due to the work of Maxwell, Boltzmann, Gibbs [15] and others. Entropy has then been increasingly popular in computer science and information theory, particularly through the paper of Shannon [16].

Maxwell, Boltzmann and Gibbs extended the notion of entropy from thermodynamics into the domain of statistical mechanics, where *macrostates* and *microstates* play an important role. The definition of entropy by Boltzmann [15] is

$$-k \sum_i P_i \log P_i$$

where the P_i are the probabilities that particle i will be in a given microstate, and all the P_i are evaluated for the same macrostate; k is the famous Boltzmann constant. Therefore, entropy in statistical mechanics denotes the uncertainty about which state the system is in.

Entropy, as defined in the work of Shannon, represents the *information content* of a message or, from the point of view of the receiver, the uncertainty about the message the sender produced, prior to its reception. It is defined as

$$-\sum_i p(i) \log p(i).$$

$p(i)$ is the probability of receiving message i and Shannon has shown that $\log p(i)$ is the only function that satisfies all requirements to measure the information or equivalently the uncertainty of a message. Later, Renyi proved the uniqueness of $\log$ as a measure of information in [17]. The unit used is the *bit* invented by John Tukey, Shannon's colleague at Bell Labs, which is the appropriate unit for entropy because of the conventional use of base-two logarithms in computing Shannon entropy. The entropy is at maximum value when all events are equiprobale, which is intuitive, because in this case the receiver has maximum uncertainty about which message he may receive next. On the other hand entropy is at minimum, if $p(i) = 1$ or $p(i) = 0$ for all i, denoting an event that occurs every time or never, so the receiver knows prior to the receipt of the message what he will get.

Entropy is nowadays heavily used for problems ranging from quantum computing to black hole physics (i.e. Tsallis entropy and many more extensions [18]). We do not use these kinds of entropy, but stick with the classical approach of Shannon and Renyi. It must be noted that despite the huge number of papers being published these days on this topic, simpler facets of entropy had still been omitted [19].

III. Model for Identification of Ontologies

The following section details at first one classical way of identifying and structuring information data with the help of a probability distribution, entropy and a distance measure. This approach is extended to encompass every kind of knowledge as to enable the identification of ontologies from it. A simple example at the end of the section illustrates the theory.

A. Classical Way

To propose a model of identification for ontologies, first let us have a look at one classical way of identifying data. There are usually various information sources available, providing data in the form of numbers from some measurements. This data normally represents information about an experiment or some kind of process to be examined, and needs to be interpreted as to learn something about the process. One way to accomplish this is to find some meaningful structure in the data that may result in new insights about the measured process, i.e. the structure represents similarities to an already known experiment, but differs in one important aspect.

There are several ways of identifying a structure on data; the usual one is to divide it into clusters that are then used to identify some sort of classification algorithms on the available data. The approach we are examining involves the concepts of a probability distribution, entropy and finally an associated distance measure. After we have acquired the data, the first step is therefore to define a matching probability distribution that *may* represent the way, how the data was generated by the process.

After a matching distribution is found, we can impose a structure on our data with the concept of entropy. One of the many uses of entropy is the notion of information content, defined by Shannon in [16] and generalized by Renyi in [20]. The meaning of entropy for our context is, that it measures the *amount* of information needed to specify an element under consideration from a given set of elements.

With this interpretation of entropy, we can introduce the concept of distance to impose a structure on the data, generated by the information sources. There are several definitions for an information-theoretic distance, like for example the Chernoff [21] and Kullback-Leibler distance (both belonging to the Ali-Silvey class of information-theoretic distance measures [22]). We haven chosen the Kullback-Leibler distance [23] because of its simplicity and computational effectiveness. It should be noted, that the Kullback-Leibler distance is not a real distance measure, at least in the metric (measure-theoretic) sense, because it violates the symmetry axiom, but is nonetheless called a distance, simply by analogy. The Kullback-Leibler (or relative entropy) measure describes the distance between a real distribution $\mathbf{p} = p_1, \dots, p_n$ in a system and an assumed one $\mathbf{q} = q_1, \dots, q_n$:

$$D(\mathbf{p}||\mathbf{q}) = \sum_i^n p_i \log \frac{p_i}{q_i}$$

Those three steps, the selection of a distribution matching the data, the use of entropy and an appropriate distance measure, enables us to structure the original data sources (see figure 1). In the next section, we extend this approach to encompass not only (information) data but also general knowledge.

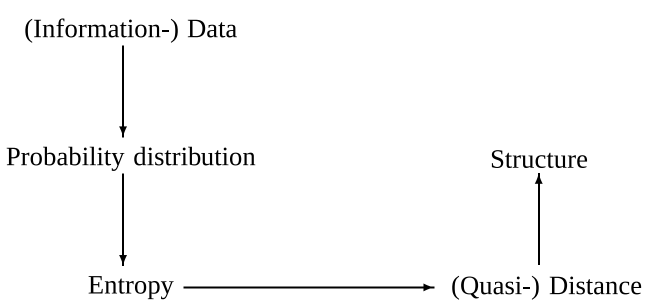

Fig. 1. Classical model of (information-) data

B. Model for identification of ontologies

In our extended model, we allow as information sources not only information data, but also any kind of knowledge. We do not make any assumptions about it, so the knowledge can be complete or incomplete, certain or uncertain and structured or unstructured. Ontologies or taxonomies, as two examples of embodied knowledge, are nowadays usually incomplete, because they are domain specific, but may be seen as complete for a particular domain. This kind of knowledge is generally viewed as certain and in some sense structured, though the structure may not be clear, i.e. the Nixon-Quaker-Republican example (Nixon Diamond due to Reiter [24]).

Like in the classical way, we have to choose at the beginning a distribution for our knowledge, which is one of the key points in our model to generate structured knowledge. In a first approach this distribution results from assessments by a given expert, which can be interpreted as the ”view of the world” of this expert and the distribution is surely different for another expert. The view taken on the domain by this expert may be incomplete due to limited knowledge and can also be in contrast to an agreed upon view hold by other experts. The distribution henceforth represents a subjective view on the knowledge, which can be interpreted as giving a weight to the concepts to represent their significance, for a particular person. If we take for example a taxonomy as already existing knowledge, then different persons may assign different distributions to the concepts in the taxonomy, depending on their view on that domain. It is also possible that the literature on this domain is incomplete or conflicting, so an approach like Renik's (see section II-A) may lead to unwanted results, for example if the information content of important concepts is near 0.5 (representing total ignorance), may be due to odd frequencies of the words in the incomplete or conflicting documents. The distribution, assigned by an expert, enables us to define an entropy, like in the classical case, and consequently a distance measure, which in turns facilitates the structuring of knowledge.

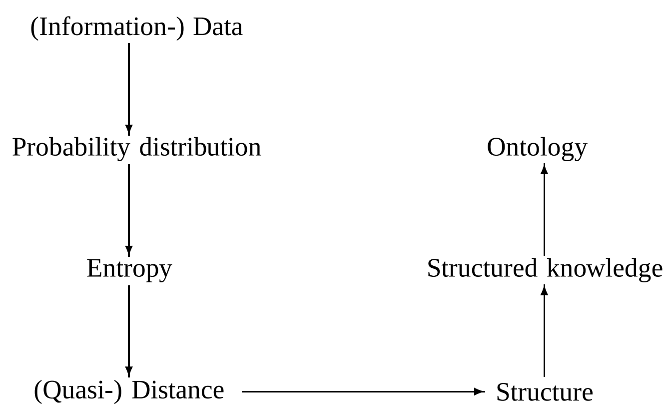

Fig. 2. Model for identification of ontologies

Our approach to structure knowledge is summarized in figure 2: Beginning from arbitrary knowledge, on which we impose in a first step a probability distribution, representing for example a subjective view of a person, we can create the

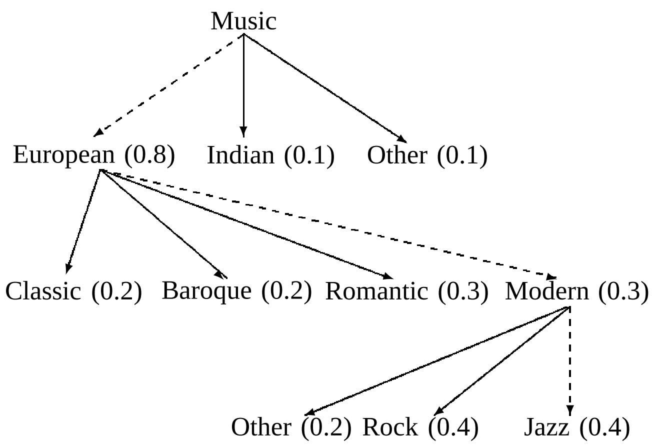

Fig. 3. Taxonomy with equal preferences for Jazz and Rock

notion of entropy. With the help of entropy we can introduce the concept of distance and hence establish a distance measure on knowledge, which means we can structure it. This in turn imparts we have shown a way to get structured knowledge but this is, by definition, an ontology.

C. Example

In figure 3 a simple example taxonomy for music styles has been constructed. Each concept has an assigned probability representing for example a preference for this concept. They have been assigned in such a manner, that they sum up to one for each hierarchy level, as it is required for a probability distribution. For our example we consider only the dashed path along the concepts *European*, *Modern* and *Jazz*.

Both taxonomies display a strong preference for European music, where Indian and other music styles are not really considered. On the second hierarchy level there are nearly equal preferences in figure 3, with a slight preference for modern and classical music. This has changed drastically in figure 4, where a new interpretation of the ontology has emerged, which focuses on the music of Bizet and Mendelsson (Romantic) and modern music. The different interpretations hold also at the last level: In figure 3 interests between rock and jazz are balanced, but in figure 4 preferences have changed in favor of jazz.

For our simple example we simplify the calculation of the Kullback-Leibler distance in such a way, that we will use as probability distribution **p**, respectively **q**, for the hierarchy levels of the ontology the probability p of the relevant concept (those along the dashed path) and the sum of the rest of the concepts $1-p$ at this level. Let us further define the index P, respectively Q, for all the probabilities of the taxonomy represented by figure 3, respectively figure 4.

The Kullback-Leiber distance

$$D(\mathbf{p}||\mathbf{q}) = \sum_i p_i \log \frac{p_i}{q_i}$$

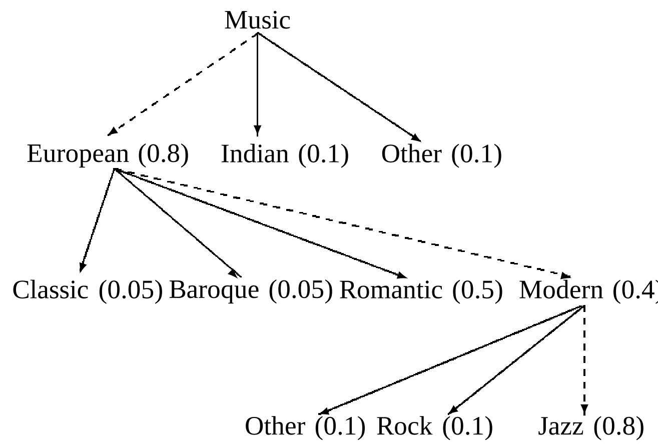

Fig. 4. Taxonomy with preferential focus on Jazz

between the European concepts $D(European_P||European_Q)$ is obviously 0, because they have the same preference in both taxonomies, e.g. no change regarding this aspect of music. $D(Modern_P||Modern_Q) = -0.12 + 0.15 = 0.03$ shows a slight distance between both concepts, because in the Q-taxonomy modern music is slighty favored compared to the P-taxonomy. In our simple example we do not pay attention to the changes to the other probabilities at this hierarchy level. Now, $D(Jazz_P||Jazz_Q) = -0.4 + 0.95 = 0.55$ is a rather large divergence between the two concepts, because we have quite a drastical change in the preferences regarding rock, jazz. The total distance between the P- and Q-taxonomy for the dashed path is the sum of the distances on each hierarchy level: $D = 0 + 0.03 + 0.55 = 0.58$.

This means, that with the same taxonomy, e.g. the same concepts and structure of the taxonomy, we have different ontologies depending on the preferences imposed by the probability distribution on the taxonomy. The distance $D = 0.58$ henceforth summarizes the divergence of the preferences, albeit only for the dashed path. In our example, the distance has been formalized through the Kullback-Leibler measure, though more advanced ones are surely possible, that may take into account some conditional or even nonlinear dependencies.

IV. Application

In a real world application the distance measure, and the resulting structure, can be used to calculate the average distance between two possible worlds represented via ontologies or taxonomies with different probability distributions. The distribution represents the interpretation, may be in regard to some goals, of the world, represented by the ontology, of a specific person. So, if an average user has made up a simple represention of his domain of interest, like the one in figure 3, he can easily compare it to another persons view on this domain or only a part thereof, for example only the dashed path in figure 4. The complexity of a comparison from the root node to the compared concept (i.e. from Music to Jazz in figure 3) is rather low. For each concept we have to calculate the difference between the two probabilities and the complement thereof. So for n nodes we have $2n$ operations hence the complexity class is $O(n)$.

We can also compare taxonomies (or ontologies) whose distributions have been calculated, may be automatically, over some text corpus. The distance measure will then show the similarity between a users world and that of the taxonomy representing the world of a particular document. An example is a user, who has made up a taxonomy of music and wants to find similar, according to his point of view, documents via a search engine.

V. Ontologies and trust

Before outlining possible links between ontologies and trust we first provide an overview of what trust implies. To build up trust some sort of modeling is required. Models for trust range from practical ones that are already implemented to very complex ones that are under investigation. Recent references are the book edited by Katsikas et al. [25], the paper of Marchiori [26] or the documentation available from the FIPA website. The European FP6 "TrustCom" project [27] on trust management provides also relevant information on the present state of the art. Current solutions for implemented trust models are built for so-called trusted environments when simple mechanisms do exist. A first evaluation of these models demonstrates that they are composed most of the time of the following three layers:

1) An organizational layer, usually the model along which a company is organized
2) A semantic layer, which is simply a role based access model
3) A technical layer amounting to the operating system in use

Nowadays, new requirements arise from the complexity and the dynamics of our recent understanding of what security, validation and trust are about [28]. A general feature is that any decision implicitly relies on trust, whether it is made by humans or by technological means. Examples are found for instance in legal institutions, business companies or peer-to-peer computations. Trust is thus becoming a basis for security in dynamic and open systems, for instance in grid computing or in virtual organizations. Another example for the complexity of the problem is trust management for credential based access control: how can asymmetric cryptography (PKI) be used for trust management in open computing systems? To answer these challenges a series of advanced models have been designed. It is beyond the scope of this section to give an exhaustive presentation of these models, so we concentrate only on some primary features. A first idea is that party trust supplements control trust. This means that trust created by a trading party of a transaction adds to the trust created by the control mechanism of the transaction. This is different from the institutional versus dyadic trust alternative. There, a third-party institutional partner will create trust while dyadic trust is created by the trading parties themselves. These different models cover most of the options available to design trust models. However, there is plenty of room left to make more specific models:

- Computational Model of Trust where a set of partially ordered trust values and various mechanisms for calculating the trust metric are defined
- Formal Logic Models with various representations, e.g. modal logic to express sincerity, cooperation, credibility, vigilance, validity or competence are introduced
- Reputation Models where the name is self-explanatory and the methodologies cover most of the known cognitive approaches
- Socio-cognitive models of trust in terms of beliefs and goals

These models exhibit some analogy to the reputation models but they usually emphasize the implicit character encountered in most decision making system found in practical reasoning.

Trust is a prerequisite for validating any e-transaction. In [28] the legal facet of trust is outlined. This feature is central in any trust model of e-business. All the aspects of trust aim at securing e-transactions and are mandatory for any form of e-business. The link to ontology is then obvious. Indeed, a simple model for e-business is obtained when three actors are defined: the seller, the buyer and the negotiator. Each is composed of both an ontology (knowledge base) and a decision making system.

The structuring of the knowledge making up some ontology enables us to design a trust model specific for ontologies. This model is the topic of on-going work. Any of the trust models mentioned in this section do rely on a well-defined ontology, whether it is a set of partially ordered trust values or the commands of an operating system. The concept of distance on different views of the world expressed by these ontologies is a tool to define bounds on trust in a very general setting. This can be consequently introduced in both conceptual models or case study scenarios.

VI. Conclusion and further research

We have proposed a model for the identification of ontologies from arbitrary knowledge, given the well known concepts of probability distribution, entropy and distance. Entropy and distance measures are imposing a structure on our knowledge, but this is by definition an ontology. This allows us to identify differences between two or more possible worlds who share the same concepts but give those a different weight representing different interpretations of the world (example in figure 3 and figure 4). The calculation of this divergence, represented through the distance, is rather cheap and thus allows for a fast comparison between a lot of possible worlds.

A first concluding remark is that this work may look like duplicating methodologies used for instance in numerical classification or in social networks. We are simply using these techniques to illustrate in simple cases that we may identify ontologies. Moreover, we can do that efficiently since we have only to compute a distance, not the entropy of a system.

Further research has to be done to evaluate other distance measures, for example those relying on conditional probabilities. These may be mutual information, conditional mutual information and so on. Of importance is also the investigation in a generalization of Shannon entropy, namely Renyi information [29]. This allows for superadditive or subadditive probability measures (i.e. $\sum_i p_i \geq 1$ or $\sum_i p_i \leq 1$), which may be helpful for our application field. They introduce in the case of superadditive measures some redundancy, as it could be done with the concepts in our taxonomy. We can change for instance the weight for *Indian* to 0.2, thereby producing the "unusual" equation $\sum_i p_i = 1.1$, refering to some commonality between European and Indian music. Subadditive measures on the other hand represent some kind of conflict. For futher information on this topic refer to [30].

Classification of information and the evaluation of degrees of belief is another area of future work because the calculation of the distance is a rather cheap operation compared to the usual complex algorithms used in this domain. For instance the determination of the degree of belief in non-monotonic reasoning is done either through a complex rule system [31] or through algorithms computing the entropy of a system, which have to consider a huge number of different states.

Another possibility is to build some sort of expert system with taxonomies or ontologies as knowledge base. A probability distribution can be assigned to the knowledge bases and from those different kind of beliefs may be calculated with the help of a distance measure and subsequently be used in the evaluation process. The advantage of such an approach is that we can calculate those degrees of belief very fast and thus may vary the probability distribution many times in order to get an optimal result for our expert system. A last remark is that although this approach may lead to classification trees reminiscent of other domains, it will be set in a theoretical framework where dynamics is introduced into the Fisher information metric [32].

REFERENCES

[1] M. Gell-Mann, "What is complexity," *Complexity*, vol. 1, no. 1, pp. 16 – 19, 1995.

[2] T. Gruber, "Towards principles for the design of ontologies used for knowledge sharing," *Knowledge Acquisition*, pp. 199 – 220, 1993.

[3] T. Berners-Lee, J. Hendler, and O. Lassila, "The Semantic Web," *Scientific American*, 05 2001.

[4] J. Calmet and A. Daemi, "From Entropy to Ontology," in *Cybernetics and systems 2004 - AT2AI-4: From Agent Theory to Agent Implementation*, R. Trappl, Ed., vol. 2, 2004, pp. 547 – 551.

[5] H. Reichenbach, *Theory of Probability*. University of California Press, 1949.

[6] F. Bacchus, A. Grove, J. Y. Halpern, and D. Koller, "From statistics to beliefs," in *National Conference on Artificial Intelligence*, 1992, pp. 602 – 608.

[7] J. Pearl, "Probabilistic semantics for nonmonotonic reasoning: A survey," in *Readings in Uncertain Reasoning*, G. Shafer and J. Pearl, Eds., 1990, pp. 699 – 710.

[8] M. Goldszmidt, P. Morris, and J. Pearl, "A maximum entropy approach to nonmonotonic reasoning," *IEEE Transactions on Pattern Analysis and Machine Intelligence*, vol. 15, no. 3, March 1993.

[9] W. N. Fancis and H. Kucera, "Frequency Analysis of English Usage: Lexicon and Grammar," *Houghton Mifflin*, 1982.

[10] P. Resnik, "Using information content to evaluate semantic similarity in a taxonomy," in *Proceedings of IJCAI*, 1995, pp. 448–453.

[11] A. Hotho, A. Maedche, and S. Staab, "Ontology-based Text Document Clustering," *Künstliche Intelligenz*, pp. 48 – 54, 04 2002.

[12] K. S. Jones and P. Willett, Eds., *Readings in Information Retrieval*. Morgan Kaufmann, 1997.

[13] I. K. Sethi, "Document representation and IR models," http://www.cse.secs.oakland.edu/isethi/IR/Coursenotes/Notes1.pdf, 1999.

[14] J. C. Maxwell, *Theory of Heat*. Dover; Reprint, 2001.

[15] R. C. Tolman, *The principles of statistical mechanics*. Dover, 1979.

[16] C. E. Shannon, "A mathematical theory of communication," *Bell System Technical Journal*, vol. 27, pp. 379 – 423, 623 – 656, July and October 1948.

[17] A. Renyi, *Probability theory*. Springer Verlag, 1970.

[18] A. G. Bashkirov, "On the Renyi entropy, Boltzmann Principle, Levy and Power-law distributions and Renyi parameter," *eprint arXiv:cond-mat/0211685*, 2003.

[19] J. Calmet and X. Calmet, "Metric on a statistical space-time," *Submitted*, 2004.

[20] A. Renyi, "Some fundamental questions of information theory," in *Selected papers of Alfred Renyi*. Hungarian Academy of Sciences, 1976, vol. 2, pp. 526 – 552.

[21] H. Chernoff, "Measure of asymptotic efficiency for tests of a hypothesis based on the sum of observations," *Annal Math. Stat.*, vol. 23, pp. 493 – 507, 1952.

[22] S. M. Ali and S. D. Silvey, "A general class of coefficients of divergence of one distribution from another," *Journal of the Royal Statistical Society. Series B (Methodological)*, vol. 28, no. 1, pp. 131 – 142, 1966.

[23] S. Kullback and R. Leibler, "On information and sufficiency," *Annal Math. Stat.*, vol. 22, pp. 79 – 86, 1951.

[24] R. Reiter, "A logic for default reasoning," *Artificial Intelligence*, vol. 13, pp. 81 – 132, 1980.

[25] S. Katsikas, J. Lopez, and G. Pernul, Eds., *Trust and Privacy in Digital Business*. Springer-Verlag, 2004.

[26] M. Marchiori, "W5: The five w's of the world wide web," ser. Second International Conference on Trust Management. Springer-Verlag, 2004.

[27] (2004) Trustcom. EU FP6. [Online]. Available: http://www.eu-trustcom.com

[28] J. Calmet and R. Endsuleit, "An agent framework for legal validation of e-transactions in the law of electronic agents," ser. LEA 04 Workshop, C. Cevenini and G. Edizioni, Eds. Springer-Verlag, 2004, pp. 181 – 185.

[29] A. Renyi, "On measures of entropy and information," in *Selected papers of Alfred Renyi*. Hungarian Academy of Sciences, 1976, vol. 2, pp. 565 – 580.

[30] G. J. Klir and M. J. Wierman, *Uncertainty-Based Information*, ser. Studies in fuzziness and soft computing. Springer Verlag, 1998.

[31] J. L. Pollock, *Nomic Probabilities and the Foundations of Induction*. Oxford University Press, 1990.

[32] X. Calmet and J. Calmet, "Dynamics of the fisher information metric," *Submitted*, 2004.