A Formalization of Logical Imaging for Information Retrieval using Quantum Theory

First author
Department
University
Country
email

Second author
Department
University
Country
email

Third author
Department
University
Country
email

***Abstract*—In this paper we introduce a formalization of Logical Imaging applied to IR in terms of Quantum Theory through the use of an analogy between states of a quantum system and terms in text documents. Our formalization relies upon the Schrödinger Picture, creating an analogy between the dynamics of a physical system and the kinematics of probabilities generated by Logical Imaging. By using Quantum Theory, it is possible to model more precisely contextual information in a seamless and principled fashion within the Logical Imaging process. While further work is needed to empirically validate this, the foundations for doing so are provided.**

I. Introduction

In the last few years there have been several attempts to model classical systems using Quantum Theory (QT). For example, QT inspired models have been developed for the modeling of cognitive processes, such as concept formation and concepts combination [1], the modeling of semantics [6], [7], [5], [8] and the modeling of Information Retrieval (IR) processes and techniques [16], [3], [24]. The appeal of using QT in the development of such models is because it acts as a *trade d'union* between logics, probability and geometries to provide a unified point of view, while also naturally modeling the contextual behaviour of complex systems [17].

In this paper, we propose a way of formalizing the Logical Imaging (LI) technique for IR [12] within a framework based on QT [24]. By reformulating the LI within a QT framework it is expected that new instruments for improving the technique will be possible by taking advantage of the appeal of QT inspired models. To this aim, we show how LI can be successfully mapped into the QT framework, where our formalization is based on the metaphor of representing terms as states of a quantum system and documents as mixtures of such states. The technique for updating probabilities associated with the states of a system, i.e the Logical Imaging, is then modeled as a process that evolves the system by collapsing some of the states into the states that strictly belong to each document. This update relies upon the *Schrödinger Picture*, which depicts the evolution over time of the states of a quantum system. By placing LI within a QT framework the mathematical basis is provided to capture, model and use the contextual information associated with a term in order to understand its meaning in a specific context. This is unlike the original LI model proposed by Crestani et al which lacks such contextual precision.

The remainder of this paper is structured as follows: in section II, we briefly introduce the LI technique, then in section II-A we show how LI has been adapted to the IR problem by Crestani et al. In section III, we propose our formalization of LI in a framework based on QT. This requires the introduction of *the kinematics operator* which captures the *dynamic flow of probabilities* of transfer triggered by the LI process (section IV). In section V, we discuss the proposed formalism, illustrating how the metaphor undertaken in our approach differs from the one used in some other QT inspired models. Finally, the paper concludes with section VI where directions of future works are examined.

II. Logical Imaging

The Logical Uncertainty Principle (LUP) [22] introduced a new way of thinking about relevance: in fact, van Rijsbergen proposed evaluating the probability of relevance of a document for a given query[1], namely $P(R|q,d)$, using the probability of a conditional, $P(d \rightarrow q)$ [23]. Such a probability could be evaluated by a simple conditionalization $P(q|d)$ [15] but as Lewis showed, such a measure can take on at most only four different values [18], [19]. To overcome this limitation, van Rijsbergen suggested evaluating conditionalization by LI, using possible worlds. Lewis assumed that there are only a finite number of possible worlds[2]. Moreover, he considered a probability distribution over the class of possible worlds: each world W has a probability $P(W)$ and these probabilities sum to 1. For each world W and each proposition y there is a world W_y that is the most similar world to W where y results true. Then a probability function, the image of P, can be defined over y: we denote this as P_y. This function is defined by setting $P_y(W)$ for all worlds W equal to the sum of $P(W')$ for all worlds W' such that W'_y is identical to W. This means that the image of a probability function can be computed by moving the original probability of each world W' to W'_y. Essentially, LI revises the probability associated with a proposition y by means of the *minimal revision* to make y accepted. This notion of minimal revision, or minimal extension, is in accordance with what van Rijsbergen proposes in LUP: the truth value

[1]The reader can refer to [11] for a good survey of Probabilistic Models.

[2]This assumption is made only for mathematical simplicity and it can be removed if necessary, as is shown in [14].

of the conditional $y \rightarrow x$ in a world W is equivalent to the truth value of the consequent x in the closest world W_y to W where the antecedent y is true. Thus, the implication $y \rightarrow x$ is true at W if and only if x is true at W_y. Let $W(y)$ be a truth evaluation function which computes the truth value of a proposition y in the context of a world: $W(y)$ equals 1 if y is true at W, 0 otherwise. Let $W_y(x)$ be an extension of the previous function, which evaluates to either one, if the sentence x is true, or zero if false in world W_y:

$$W_y(x) = \begin{cases} 1 & \text{if } x \text{ is true at } W_y \\ 0 & \text{otherwise} \end{cases}$$

Thus, we can write $W(y \rightarrow x) = W_y(x)$. We are interested in the probability of proposition y: this can be computed by summing the probabilities of the worlds where the proposition is true; mathematically we have $P(y) = \sum_W P(W)W(y)$. Successively, we have to derive a new probability distribution, P_y from P, such that the probability associated with every world W is transferred to its most similar (closest) world W_y where y is true. This new probability distribution is

$$P_y(W') = \sum_W P(W)I(W', W)$$

where $I(W', W)$ assumes value 1 if $W' = W_y$, zero otherwise. In [19] it is illustrated how the probability of the conditional is the probability of the consequent after Imaging on the antecedent, $P(y \rightarrow x) = P_y(x)$. In the following section we present the technique for IR based on LI as introduced by Crestani et al.

A. Logical Imaging in IR

While LI provides an intuitive and novel approach to estimate the relevance of a document given a query, there have only been a few attempts at using LI in IR. And this has been performed in two ways; viewing documents as possible worlds [2], [21] or viewing terms as possible worlds[3] [10], [12], [13]. Here we focus on the later, and more empirically successful approach to LI, where terms are considered as objects of the Possible World Semantic.

Under this view, the set of possible worlds is represented by the set of terms T of a collection D of documents. Each document d of D is represented using terms belonging to T: then a document can either be true or not true in the context of a world. Let us assume a probability distribution P on the set T; the sum of each probability $P(t)$ associated with the term t has to sum to 1, such that $\sum_{t \in T} P(t) = 1$. Now, in order to evaluate the probability of conditionalization, namely $P(d \rightarrow q)$, we compute LI on d over all possible terms t in T, such that:

$$P(d \rightarrow q) = P_d(q) = \sum_{t \in T} P(t) t_d(q)$$

where $t_d(q)$ is the truth function which returns 1 if and only if q is true at t_d, 0 otherwise, and t_d is the closest term to t for which d is true. This process of LI on d causes a transfer of probabilities from terms not occurring in the document d to terms occurring in d. In order to apply LI a term–term similarity function is required. Such a function establishes which of the possible worlds has to be the target of the movement of probabilities, namely the closest world W_d to W where d is true. This directs the transfer of probabilities from terms absent in the document to terms that are present. The choice of a similarity measure will play a crucial role in the transfer of probabilities, which will invariably affect the empirical effectiveness of the LI IR model.

In [10], [12], [13], the similarity function used was the Expected Mutual Information Measure ($EMIM$). However, this function has some drawbacks: in particular, it only partially accounts for the context of terms. This is because the $EMIM$ value of a pair of terms is defined over the whole collection. The measure then does not take into account the local (i.e. at document level) relationship between two terms, encoding a measure of the global (i.e. at collection level) interaction between the terms instead. For example, in a collection containing, in similar amount, documents related to sport and to nature, it is likely having similar $EMIM$ values between the pairs *(bat, cricket)* and *(bat, night)*: in the first case the association reveals the sport sense of *bat*, while in the second example *bat* refers to the animal sense. Performing LI with $EMIM$ does not fully account for the context surrounding a term. While different similarity functions could be used, formalizing LI in terms of QT provides new directions to incorporate more precise localized contextual information seamlessly, as opposed to the ad hoc incorporation of more sophisticated functions within the original LI IR model.

III. LI FORMALIZED IN A QT FRAMEWORK

Let us assume a set T of cardinality $\|T\| = k$ representing the terms extracted from the collection D. We represent the terms in T as normalized vectors of a geometric space of dimension n: thus, each term $t_i \in T$ is represented by the vector $|\underline{t}_i\rangle$.[4] Let us consider a probability distribution P over

[3] We assume the reader is familiar with this work.

[4] In this paper, vectors and matrices are written in according to the Dirac notation, widely used in QT literature: thus a vector $\mathbf{x}$ correspond to $|\underline{x}\rangle$ and the matrix $\mathbf{y} \cdot \mathbf{y^T}$ corresponds to $\left|\underline{y}\right\rangle \left\langle\underline{y}\right|$. For an introduction to Dirac notation, the reader is referred to [20].

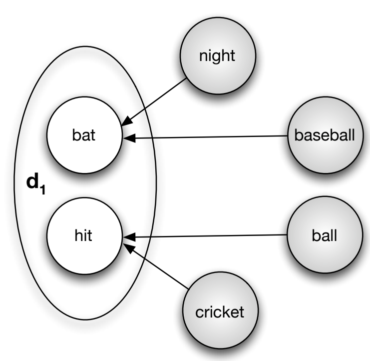

Fig. 1. Graphical interpretation of the probability kinematics induced by the Imaging process: the probabilities flow from terms outside the document (grey circles) to terms inside the document (white circles).

the set T: the distribution associates a probability $\alpha_i \in [0,1]$ with each term $t_i \in T$, such that $\sum_{i=1}^{k} \alpha_i = 1$. Since each vector $|\underline{t}_i\rangle$ corresponds to a term in T, the probability of the corresponding term is associated with each vector. If we assume that each term t_i, and the vector $|\underline{t}_i\rangle$, represents a possible world w_i, then T corresponds to the set of all possible worlds, denoted by W. Furthermore, P represents a probability distribution over W.

In order to compute $d \to q$ by LI, we need to test the document d in each possible world. A term is a d–world, if, and only if, it is present in document d. Thus, we have to transfer the probability associated with each not d-term to the closest d-term. At the end of this process, we obtain a set W_d whose elements are all the worlds (terms) where d is true. Each element of W_d is associated with a probability distribution that is given by the movements of probabilities from not d–worlds to the closest d–worlds. We can now compute the density operator ρ'_d associated with a linear combination of d–worlds, known as a mixture. Each component of the density operator is scaled by the probability of the d–world. Formally, ρ'_d is given by

$$\rho'_d = \sum_{t_i \in W_d} \alpha'_i (\sum_{j=1}^{n} \lambda_{i,j}{}^2 \left|\underline{e}_j\right\rangle \left\langle \underline{e}_j\right|)$$

where α'_i is the sum of the probability α_i with the probabilities of the not d–worlds that move to the world w_i represented by the term t_i. In the above equation, $\lambda_{i,j}$ is the j^{th} component of the normalized vector $|\underline{t}_i\rangle$ [5] and $\left|\underline{e}_j\right\rangle$ represents the j-th vector of the canonical basis.

We can calculate the projector which represents the subspace spanned by a document d by $\mathbf{P}_d = \bigcup_{t_i \in d} \left|\underline{t_i}\right\rangle \left\langle \underline{t_i}\right|$: similarly, the projector associated with a query q is given by $\mathbf{P}_q = \bigcup_{t_i \in q} \left|\underline{t_i}\right\rangle \left\langle \underline{t_i}\right|$. Thus, the probability $P(d \to q)$ calculated by LI in the proposed QT framework is given by the probability of the subspace $[\![\mathbf{P}_d \to \mathbf{P}_q]\!]$ calculated by the trace operation $tr(\rho'_d \mathbf{P}_R)$, where $\mathbf{P}_R$[6] is the projector associated with $[\![\mathbf{P}_d \to \mathbf{P}_q]\!]$. One may argue that ρ'_d is not a density operator but just a positive self–adjoint (Hermitian) operator whose trace is equal to one: the interested reader can refer to appendix B for the demonstration that ρ'_d respects the definition of density operator.

IV. The Kinematic Operator

Given the proposed QT interpretation of LI, it is possible to define a linear transformation of the matrix of probabilities ρ_d in order to generate the movement of probabilities from one not d–world to its closest d–world. This transformation is referred to as kinematics operator denoted by $\mathbf{K}$ and is analogous to the approach taken by Schödinger in order to address the evolution of a quantum system.

In the previous section, we computed the matrix of probabilities after the LI process by simply calculating ρ'_d as a linear combination of new probabilities α'_i multiplied by the relative projectors $\mathbf{P_i}$. In the following, we show how the LI process can be represented mathematically and how to compute ρ'_d using a linear transformation applied to ρ_d[7]. We build a $k \times k$ matrix $\mathbf{K}$ by filling it with ones in the position (i,i) if the document d is true in the world w_i – representing the term t_i – and 0 in the other positions. Corresponding to each row i where (i,i) is equal to zero we set to 1 the entry (i,j) if the world (term) w_j is the closest world (term) to w_i where d is true. In representing $\mathbf{K}$, the non-diagonal entries which have value 1 encode the movement of probabilities from the term i to the term j. Thus, the density operator ρ'_d after LI is expressed by the transformation $\tau : \rho_d \to^{\tau} \rho'_d$, where the transformation is represented by $\rho'_d = \mathbf{K}^T \rho_d \mathbf{K}$.

Let us analyze the proposed transformation $\mathbf{K}^T \rho_d \mathbf{K}$. In particular, $\rho_d \mathbf{K}$ will generate a matrix whose diagonal entries are the original probabilities for the d–worlds and 0 otherwise. The probabilities associated with the not d–worlds are not lost but moved to the position (i,j) where i is the index associated with the term t_i representing a not d–world and j the one associated with the closest d–world to t_i, namely t_j. After applying the transformation $\rho_d \mathbf{K}$ we obtain a matrix whose columns contain the probabilities that are associated (after LI) with term index of each column. Note that $\mathbf{K}$ encodes in an intuitive way the information about the source of the probabilities involved in the LI process. It allows us to understand which terms contribute to incrementing the probability associated with the d–term t_j by selecting column j of $\mathbf{K}$ and considering its non–zero components: in correspondence with such components there would be the not d–terms which move their probabilities to t_j. In order to complete the transformation, $\mathbf{K}^T$ has to be applied, obtaining the new density operator ρ'_d on whose diagonal entries the probabilities of each term in the term–space after LI are encoded. That is $\rho'_d(i,i) = \sum_{r=1}^{k} \mathbf{K}^T_{i,r} (\sum_{s=1}^{k} \rho_{d\,i,s} \mathbf{K}_{s,i})$. Let us consider the following example in order to understand better the behaviour of the proposed operator in the description of the probability kinematics induced by LI. Suppose a document d_1 is represented by terms *"bat"* and *"hit"*; the probabilities of the terms in the information space are encoded in the diagonal entries of the density operator ρ_{d_1} (Fig. 2). Given that the most similar term to *baseball* and to *night* is *bat*, while the most similar one to *cricket* and to *ball* is *hit*, we can encode this information in the kinematics operator $\mathbf{K}$ (Fig. 3). By the application of $\tau : \rho_{d_1} \to^{\tau} \rho'_{d_1}$ we obtain the operator ρ'_{d_1} as shown in Fig. 4.

$$\rho_{d_1} = \begin{bmatrix} bat & ball & night & cricket & hit & baseball \\ 0.2 & 0 & 0 & 0 & 0 & 0 \\ 0 & 0.1 & 0 & 0 & 0 & 0 \\ 0 & 0 & 0.05 & 0 & 0 & 0 \\ 0 & 0 & 0 & 0.2 & 0 & 0 \\ 0 & 0 & 0 & 0 & 0.3 & 0 \\ 0 & 0 & 0 & 0 & 0 & 0.15 \end{bmatrix}$$

Fig. 2. The density operator ρ_{d_1}

[5]Since it is a unit vector, we can be sure that the square of its components sums to 1.

[6]Refer to appendix A for the calculation of $\mathbf{P}_R$.

[7]The density operator associated with a document before the LI process is given by $\rho_d = \sum_{t_i \in W_d} \alpha_i (\sum_{k=1}^{n} \lambda_{i,k}{}^2 \left|\underline{e}_k\right\rangle \left\langle \underline{e}_k\right|)$, where α_i is the original probability of term t_i.

$$\mathbf{K} = \begin{bmatrix} \mathbf{1} & 0 & 0 & 0 & 0 & 0 \\ 0 & 0 & 0 & 0 & 1 & 0 \\ 1 & 0 & 0 & 0 & 0 & 0 \\ 0 & 0 & 0 & 0 & 1 & 0 \\ 0 & 0 & 0 & 0 & \mathbf{1} & 0 \\ 1 & 0 & 0 & 0 & 0 & 0 \end{bmatrix}$$

Fig. 3. The Kinematics operator $\mathbf{K}$

$$\rho\prime_{d_1} = \mathbf{K}^T(\rho_{d_1}\mathbf{K}) = \mathbf{K}^T \begin{bmatrix} 0.2 & 0 & 0 & 0 & 0 & 0 \\ 0 & 0 & 0 & 0 & 0.1 & 0 \\ 0.05 & 0 & 0 & 0 & 0 & 0 \\ 0 & 0 & 0 & 0 & 0.2 & 0 \\ 0 & 0 & 0 & 0 & 0.3 & 0 \\ 0.15 & 0 & 0 & 0 & 0 & 0 \end{bmatrix} =$$

$$= \begin{bmatrix} 0.4 & 0 & 0 & 0 & 0 & 0 \\ 0 & 0 & 0 & 0 & 0 & 0 \\ 0 & 0 & 0 & 0 & 0 & 0 \\ 0 & 0 & 0 & 0 & 0 & 0 \\ 0 & 0 & 0 & 0 & 0.6 & 0 \\ 0 & 0 & 0 & 0 & 0 & 0 \end{bmatrix}$$

Fig. 4. The density operator $\rho\prime_{d_1}$ obtained after LI

A. Analogy to the Schrödinger Picture

It is interesting to note that the proposed kinematics operator plays a similar role to the evolution operator in the Schrödinger Picture. In fact, the time evolution of a density operator $\mathbf{D}$ representing a state is given by the Schrödinger equation, stating $i\hbar\frac{d\mathbf{D}(t)}{dt} = \mathbf{HD}(t)$, from which an evolution operator $\mathbf{U}(t)$ can be defined as $\mathbf{U}(t) = exp(-i\frac{\mathbf{H}t}{\hbar})$. The *evolution operator* is the counterpart of the introduced *kinematic operator* and provides the analogy between the dynamics of a quantum system and the transferring of probabilities in the LI process.

However, there is one main difference between the two operators. The evolution operator is unitary: $\mathbf{U}(t)\mathbf{U}^{\dagger}(t) = \mathbf{U}^{\dagger}(t)\mathbf{U}(t) = I$ [8]. In particular, if the entries of $\mathbf{U}$ are restricted to the real numbers, then $\mathbf{U}^{\dagger}$ becomes $\mathbf{U^T}$, and since $\mathbf{U}$ is an orthogonal matrix, then the equality $\mathbf{U^T} = \mathbf{U^{-1}}$ is valid. Moreover, the operator $\mathbf{U}$ conserves the inner product, $\langle \mathbf{U}(t)\beta | \mathbf{U}(t)\alpha \rangle = \langle \beta | \alpha \rangle$. Beltrametti and Cassinelli [4] write that "the dynamical evolution expressed by $\mathbf{D}(t_2) = \mathbf{U}(t_2)\mathbf{D}(t_1)\mathbf{U}^{\dagger}(t_2)$ [9] preserves the convex structure of states: if $\mathbf{D}(t_1)$ is a mixture, say $\mathbf{D}(t_1) = w_1\mathbf{D}'(t_1) + w_2\mathbf{D}''(t_1)$, then $\mathbf{D}(t_2) = w_1\mathbf{D}'(t_2) + w_2\mathbf{D}''(t_2)$. [...] The preservation of the convex structure of states is, from the physical point of view, a rather general requirement; nevertheless, there are some concrete situations in which dynamical evolutions can occur that do not preserve convexity." This is the case of the kinematics operator, because we do not want to preserve the convex structure of the states: we want the structure to change instead.

[8] $\mathbf{U}^{\dagger}$ is the self–adjoint (Hermitian) operator to $\mathbf{U}$.

[9] In their original work Beltrametti and Cassinelli wrote $\mathbf{D}(t_2) = \mathbf{U}_{t_2-t_1}\mathbf{D}(t_1)\mathbf{U^{-1}}_{t_2-t_1}$, but we prefer to maintain our formulation of such expression in order to do not generate confusion using different notations among this paper.

V. Related QIR Models

Our formalization of LI is based on the metaphor of representing terms as states of a quantum system and documents as mixtures of such states. LI is then modeled as a process that evolves the system by collapsing some of the states into the states that strictly belong to each document, e.g. the terms in the document under consideration. Of the QT inspired models the most similar approaches to ours have been put forward by [6] and [7]. In [6], terms are similarly represented by states, but in a wider sense. In fact, the authors suggest interpreting a word as a *massively entangled state* which would collapse into a simpler state once a particular meaning related to the word has been selected. This suggests that at the time of collapse the current context would influence the target state into which the superposition is collapsed. The same approach is taken in [7] in which words belonging to a semantic space are associated with quantum particles. Specifically, when the context of a word is not considered, it is represented by a superposition of states, each one of which is associated instead with one of the particular meanings of the term. Since a density operator is associated with each of the basis states (i.e. one of the several possible senses of a term), the single eigenvector associated with such density operator represents a source of context–sensitive information between the associations carried by the semantic vector under consideration. Thus, all the meanings associated with a word give rise to a complex density operator which can be constructed as a linear combination of simpler density matrices associated with each of the meaning. Moreover, a probability can be ascribed to each meaning of a word; formally $\rho_t = \alpha_1\rho_1 + \ldots + \alpha_m\rho_m$ represents the density operator corresponding to term t in an uncertain context as linear combination of its m meanings $\rho_1, \ldots, \rho_m$ weighted by the probability of the related meaning. Conversely, in our approach we associate a document to a density operator constructed as a linear combination of the density matrices associated with the terms in the information space. Successively, the collapse of the density operator into a new representation by means of the application of the kinematics operator addresses the belief revision required by LI. This is similar to the way the expression of the context declaration (i.e. the specific meaning of a term) is modeled by the collapse of the density operator [7]. Consequently the way in which we have formalized LI in the QT is able to take into account the specific context of the terms, which enables the more precise incorporation of context within the model; and paves the way for developed a context based LI model, which we shall explore in further work.

VI. Conclusion

The main contribution of this paper is the formalization of the LI IR model within a QT framework. The benefit of this approach is that contextual information can be injected in the LI technique. We have illustrated how to model the transfers of probabilities involved in the LI process by means of the kinematic operator. The kinematics operator, based on the Schrödinger Picture, will be the subject of future research.

In particular, we are interested in investigating the utility of a particular tessellation of the geometric space, namely the Voronoi diagrams (VoDs). In fact, we believe that VoDs can effectively model the underling contextual evidence intrinsic in the information space used in our QT framework. Further, it can be potentially be used to guide the updating of probability caused by the application of the kinematics operator. In future work, we will explore two main directions; how VoDs can be incorporated to produce context based LI models, and how well these QT inspired models perform against the original LI IR models and state of the art IR models.

ACKNOWLEDGMENT

[Blinded for peer review].

REFERENCES

[1] D. Aerts and L. Gabora. A Theory of Concepts and their Combinations I. *Kybernetes*, 34:151–175, 2005.

[2] G. Amati and S. Kerpedjev. An Information Retrieval Model: Implementation and experiments. Technical Report 5B04892, Fondazione Ugo Bordoni, Roma, IT, 1992.

[3] S. Arafat and C. J. van Rijsbergen. Quantum Theory and the Nature of Search. In *Proc. of the 1st AAAI Quantum Interaction Symposium*, pages 114–121, Stanford, CA, USA, 2007.

[4] E. G. Beltrametti and G. Cassinelli. *The Logic of Quantum Mechanics*. Addison–Wesley, Reading, MA, USA, 1981.

[5] P. Bruza and R. Cole. Quantum Logic of Semantic Space: An Exploratory Investigation of Context Effects in Practical Reasoning. In *We Will Show Them! Essays in Honour of Dov Gabbay*, volume 1, pages 339–362. College Publications, 2005.

[6] P. Bruza, K. Kitto, D. Nelson, and K. McEvoy. Entangling Words and Meaning. In *Proc. of the 2nd AAAI Quantum Interaction Symposium*, pages 118–124, Oxford, UK, 2008.

[7] P. Bruza and J. Woods. Quantum Collapse in Semantic Space: Interpreting Natural Language Argumentation. In *Proc. of the 2nd AAAI Quantum Interaction Symposium*, pages 141–147, Oxford, UK, 2008.

[8] S. Clark, B. Coecke, and M. Sadrzadeh. A Compositional Distributional Model of Meaning. In *Proc. of the 2nd AAAI Quantum Interaction Symposium*, pages 133–140, Oxford, UK, 2008.

[9] D. W. Cohen. *An Introduction to Hilbert Space and Quantum Logic*. Springer–Verlag, 1989.

[10] F. Crestani, M. Lalmas, and C. J. van Rijsbergen, editors. *Information Retrieval: Uncertainty and Logics : Advanced Models for the Representation and Retrieval of Information*, chapter 10 – Logical Imaging and Probabilistic Information Retrieval, pages 247 – 279. Kluwer Academic Publishing, 1998.

[11] F. Crestani, M. Lalmas, C. J. van Rijsbergen, and I. Campbell. "Is this document relevant?... Probably": a survey of Probabilistic Models in Information Retrieval. *ACM Computing Surveys (CSUR)*, Jan 1998.

[12] F. Crestani and C. J. van Rijsbergen. Information Retrieval by Logical Imaging. *Journal of Documentation*, 51(1):3–17, Mar 1995.

[13] F. Crestani and C. J. van Rijsbergen. Probability Kinematics in Information Retrieval. *Proc. ACM SIGIR '95*, pages 291–299, Jan 1995.

[14] P. Gärdenfors. *Knowledge in Flux. Modeling the Dynamics of Epistemic States*. The MIT Press, Cambridge, MA, USA, 1988.

[15] W. L. Harper, R. Stalnaker, and G. Pearce, editors. *Ifs. Conditionals, Belief, Decision, Chance and Time*. Reidel, Dordrecht, NL, 1981.

[16] A. Huertas-Rosero, L. A. Azzopardi, and C.J. van Rijsbergen. Characterising Through Erasing – A Theoretical Framework for Representing Documents Inspired by Quantum Theory. In *Proc. of the 2nd AAAI Quantum Interaction Symposium*, pages 160–163, Oxford, UK, 2008.

[17] K. Kitto. Why Quantum Theory? In *Proc. of the 2nd AAAI Quantum Interaction Symposium*, pages 11–18, Oxford, UK, 2008.

[18] D. K. Lewis. *Counterfactuals*. Basil Blackwell, Oxford, UK, 1973.

[19] D. K. Lewis. Probabilities of Conditionals and Conditional Probabilities. *The Philosophical Review*, 85:297–315, 1976.

[20] L. Sadun. *Applied Linear Algebra. The Decoupling Principle*. Prentice Hall, 2001.

[21] T. Sembok and C. J. van Rijsbergen. Imaging: a Relevance Feedback Retrieval with Nearest Neighbour Clusters. In *Proc. of the BCS Colloquium in Information Retrieval*, pages 91–107, Glasgow, UK, 1993.

[22] C. J. van Rijsbergen. A new Theoretical Framework for Information Retrieval. In *Proc. ACM SIGIR '86*, Pisa, IT, Jan 1986.

[23] C. J. van Rijsbergen. Towards an Information Logic. In *Proc. ACM SIGIR '89*, pages 77–86, Cambridge, MA, USA, 1989.

[24] C. J. van Rijsbergen. *The Geometry of Information Retrieval*. CUP, 2004.

APPENDIX A

In the following we suggest how to compute the projector $\mathbf{P}_R$ associated with the subspace $[\![\mathbf{P}_d \rightarrow \mathbf{P}_q]\!]$, also known as the Subspace conditional (S-conditional)[24]. This subspace is defined by $[\![\mathbf{P}_d \rightarrow \mathbf{P}_q]\!] = \{|\underline{x}\rangle : \mathbf{P}_q\mathbf{P}_d|\underline{x}\rangle = \mathbf{P}_d|\underline{x}\rangle, |\underline{x}\rangle \in \mathbf{P}_q\}$. Following [24], the semantics of $\mathbf{P}_d \rightarrow \mathbf{P}_q$ is given by $[\![\mathbf{P}_d \rightarrow \mathbf{P}_q]\!] = [\![\mathbf{P}_d]\!]^\perp \oplus ([\![\mathbf{P}_d]\!] \cap [\![\mathbf{P}_q]\!])$, where $[\![\mathbf{P}_i]\!] \oplus [\![\mathbf{P}_j]\!]$ is the smallest subspace containing $\mathbf{P}_i$ and $\mathbf{P}_j$. Let us examine the case in which the projectors of document and query do not commute. Then, $\mathbf{P}_d \wedge \mathbf{P}_q$ is given by $\lim_{n\rightarrow\infty}(\mathbf{P}_d\mathbf{P}_q\mathbf{P}_d)^n$, and $\mathbf{P}_d \vee \mathbf{P}_q = (\mathbf{P}_d^\perp \wedge \mathbf{P}_q^\perp)^\perp$ is given by $\mathbf{I} - \lim_{n\rightarrow\infty}((\mathbf{I}-\mathbf{P}_d)(\mathbf{I}-\mathbf{P}_q)(\mathbf{I}-\mathbf{P}_d))^n$. Thus, the projector associated with the subspace $[\![\mathbf{P}_d \rightarrow \mathbf{P}_q]\!]$ is computed as $\mathbf{P}_R = \mathbf{I} - \lim_{n\rightarrow\infty}((\mathbf{I} - \mathbf{P}_d)(\mathbf{I} - \lim_{n\rightarrow\infty} \mathbf{P}_d\mathbf{P}_q\mathbf{P}_d)^n(\mathbf{I} - \mathbf{P}_d))^n$.

APPENDIX B

A density operator is in an one–to–one relationship with the states of a quantum system. These states could be pure statesor a mixture of pure states. For an infinite Hilbert space, a vector $|\varphi\rangle$ is a mixture if a pairwise orthogonal sequence $\langle\langle|\underline{b}_i\rangle\rangle\rangle$ of unit vectors[10] and a sequence $\langle\langle\lambda_i\rangle\rangle$ of real numbers such that (i) $\lambda_j \in [0, 1]$, (ii) $\sum_{i=1}^{\infty} \lambda_i = 1$ and (iii) $\varphi = \sum_{i=1}^{\infty} \lambda_i |\underline{b}_i\rangle$ exist [9]. Then, a density operator ρ is strictly defined as $\rho = \sum_{i=1}^{\infty} \lambda_i^2 |\underline{b}_i\rangle \langle\underline{b}_i|$. Since we are operating on a finite n–dimensional subspace of a Hilbert space, the sum can be bounded from $i = 1$ to $i = n$. We want to prove that $\rho'_d = \sum_{t_i \in W_d} \alpha'_i(\sum_{j=1}^{n} {\lambda_{i,j}}^2 \left|\underline{e}_j\right\rangle \left\langle\underline{e}_j\right|)$ respects the definition of density operator. Let us assume $W_d = \{t_1, t_2\}$ (it is straight forward then to generalize this demonstration from two worlds to n). Then,

$$\begin{gathered}\rho'_d = \textstyle\sum_{t_i \in W_d} \alpha'_i(\sum_{k=1}^{n} {\lambda_{i,k}}^2 |\underline{e}_k\rangle \langle\underline{e}_k|) = \\ \alpha'_1({\lambda_{1,1}}^2\mathbf{E_1} + {\lambda_{1,2}}^2\mathbf{E_2}) + \alpha'_2({\lambda_{2,1}}^2\mathbf{E_1} + {\lambda_{2,2}}^2\mathbf{E_2}) = \\ (\alpha'_1{\lambda_{1,1}}^2 + \alpha'_2{\lambda_{2,1}}^2)\mathbf{E_1} + (\alpha'_1{\lambda_{1,2}}^2 + \alpha'_2{\lambda_{2,2}}^2)\mathbf{E_2}\end{gathered}$$

Thus we have to demonstrate that the terms which multiply the $\mathbf{E_i}$ sum to 1:

$$\begin{gathered}\alpha'_1{\lambda_{1,1}}^2 + \alpha'_2{\lambda_{2,1}}^2 + \alpha'_1{\lambda_{1,2}}^2 + \alpha'_2{\lambda_{2,2}}^2 = \\ \alpha'_1({\lambda_{1,1}}^2 + {\lambda_{1,2}}^2) + \alpha'_2({\lambda_{2,1}}^2 + {\lambda_{2,2}}^2) = \alpha'_1 + \alpha'_2 = 1.\end{gathered}$$

If the above applies then ρ_d is a density operator. The probability distribution defined on the set of all terms induces a new distribution. In particular, the latter is a probability distribution over the set of pure states – *pure meanings*, which can be interpreted as non-ambiguos meanings such as meanings only pertaining to single terms– of the subspace $[\![\mathbf{T}]\!]$. In fact, we can express ρ'_d with respect to the projectors $\mathbf{E_1}, ..., \mathbf{E_n}$ for a space of dimension n: $\rho'_d = (\alpha'_1{\lambda_{1,1}}^2 + ... + \alpha'_n{\lambda_{n,1}}^2)\mathbf{E_1} + ... + (\alpha'_1{\lambda_{1,n}}^2 + ... + \alpha'_n{\lambda_{n,n}}^2)\mathbf{E_n} = p_1\mathbf{E_1} + ... + p_n\mathbf{E_n}$ where $p_1, ..., p_n$ sum to 1. Moreover, each p_i is in the range $[0, 1]$: then $p_1, ..., p_n$ is a probability distribution over the projectors $\mathbf{E_1}, ..., \mathbf{E_n}$, which represent the projectors of the canonical basis for a space of dimension n.

[10] A sequence given by vectors of the orthonormal basis.