What can Quantum Theory bring to Information Retrieval?

Benjamin Piwowarski
University of Glasgow
benjamin@bpiwowar.net

Ingo Frommholz
University of Glasgow
ingo@dcs.gla.ac.uk

Mounia Lalmas
University of Glasgow
mounia@acm.org

Keith Van Rijsbergen
University of Glasgow
keith@dcs.gla.ac.uk

ABSTRACT

The probabilistic formalism of quantum physics is said to provide a sound basis for building a principled information retrieval framework. Such a framework can be based on the notion of information need vector spaces where events, such as document relevance or observed user interactions, correspond to subspaces. As in quantum theory, a probability distribution over these subspaces is defined through weighted sets of state vectors (density operators), and used to represent the current view of the retrieval system on the user information need. Tensor spaces can be used to capture different aspects of information needs. Our evaluation shows that the framework can lead to acceptable performance in an ad-hoc retrieval task. Going beyond this, we discuss the potential of the framework for three active challenges in information retrieval, namely, interaction, novelty and diversity.

Categories and Subject Descriptors

H.3.3 [**Information Storage and Retrieval**]: Information Search and Retrieval

General Terms

Theory

Keywords

Model, Quantum Theory

1. INTRODUCTION

Many successful models for information retrieval (IR) exist, and nowadays increasing effort is put into user-oriented models. But there is still a lack of a unified theoretical framework able to address the various challenges identified in IR. Quantum physics, on the other hand, offers a probabilistic, logic and geometric formalism based on the mathematics of Hilbert spaces, to describe the behaviour of matter at (sub)atomic scales. As van Rijsbergen claims [16], this "language" of quantum mechanics can also be used to express the different geometric, probabilistic and logic-based IR models within a unified framework, while being able to consider user-oriented aspects. Driven by this motivation, research in the field of quantum-inspired IR is investigating how the mathematical framework and postulates behind quantum theory can be applied to attack the diverse challenges in IR. These efforts led to the recent proposal of an abstract quantum IR (QIR) framework that applies methods known from quantum mechanics, in particular the notion of Hilbert spaces, system states (expressed as density operators) and measurements, to interactive retrieval [14].

A first instantiation of this framework focused on the representation of queries (as density operators) and documents (as subspaces) [12]. Initial experiments addressed classical ad-hoc retrieval tasks, which cover a potential *first step* of user interaction, the submission of a query by a user. While we do not regard ad-hoc tasks as the main target of the framework, it is crucial to show that acceptable performance is achieved with it. In these experiments (performed on one test collection), various methodologies (and associated parameters) to construct the representations were explored, but none of them led to good retrieval performance.

The focus of this paper is to move a step further by first describing some of the interesting properties of this framework and refining the representation of queries and documents. This leads us, from a theoretical point of view, to investigate other concepts known from quantum physics, namely tensor products, and from a practical point of view, to extensively experiment with different parameters. The experiments carried out on several TREC collections show that the framework is now mature enough to support current IR challenges, namely, dealing with user interaction, novelty (of documents) and diversity (of documents and topics) in a principled manner. These challenges remain difficult or impractical to solve with current approaches [17, 22].

The outline of this paper is as follows. We first discuss related work (Section 2). We then introduce the quantum-based IR framework (Section 3). In Section 4, we show how to compute the document representation (Section 4.1) and construct the query representations (Section 4.2). These are evaluated in an ad-hoc retrieval scenario in Section 5. We discuss how the framework addresses interactivity, novelty and diversity in IR in Section 6. Finally, we conclude.

CIKM'10, October 25–30, 2010, Toronto, Ontario, Canada.

2. RELATED WORK

The quantum based IR framework relies on a multi-dimensional representation of documents (subspaces) and queries (densities). Multi-dimensional representations have been implicitly used in IR to handle negative feedback. [7] showed that without such a representation, contradictory results were obtained. Later, [17] found that negative feedback could be handled by describing the information need as a set of vectors. The QIR framework encompasses this approach in the sense that a query is represented as a weighted set of vectors, too.

A more explicit use of multi-dimensional representations is the work from [4] who proposed to randomly split a document into two parts, and to use a two-dimensional representation of documents to obtain a "stereoscopic" view of a document. Our approach can be thought of as a principled extension of this work, where we do not limit ourselves to two dimensions and, in addition, we rely on the probabilistic framework of quantum physics to compute the relevance of a document to a query.

Explicit multidimensional representations have also been explored. [23] showed that the cluster hypothesis still holds when representing documents as subspaces. Their methodology to build subspaces is close to ours since to represent documents they compute the subspace spanned by a set of vectors, albeit implicitly. In our work, we provide an explicit methodology to construct the subspaces. [9] also uses a subspace for representing a user's information need (the subspace where relevant document vectors should lie), and a vector representation for documents. The probability that a document is relevant to a user's information need is determined by the projection of its vector representation onto the corresponding (information need) subspace. Following quantum physics, we interchanged the role of document and user's information need[1]. This is motivated by the fact that the user's information need should be represented as a dynamic component, as advocated in e.g. [8].

In IR, our work also bears some similarity with Latent Semantic Indexing (LSI, [6]) since we use spectral analysis to extract document and query representations. However, we do not represent objects in a low-dimensional space as in LSI, but use a spectral analysis to obtain a compact representation of our document subspaces and query densities. Hyperspace Analogous to Language (HAL) spaces [3] are also closely related to our work. There, each word w is represented by a term vector where non null components correspond to words co-occurring within a small window centred around word w. Our representation of a term is inspired by this approach, but, for each word w, we use spectral analysis to summarise the information brought by the set of vectors associated with it.

In quantum IR, besides van Rijsbergen's seminal work [16], [22] experimented with a quantum inspired principle for ranking documents. Our work proposes another approach to the problem of diversity, whereby the representation itself gives a principled way to rank documents.

Outside IR, in the face detection domain, subspaces are commonly used to solve recognition problems. In [2], a face is represented by a subspace (generated from different picture vectors of the same face) and recognition involves computing the distance between a vector (representing the face to be recognised) and the subspace. In our work, we use a similar idea to generate document representations. These are represented by a subspace generated from different aspects of the document. Different from [2], a query also has a multi-dimensional representation.

[1] [9] refers to the user "context" rather than the user's information need, but both concepts are related.

3. THE QUANTUM IR FRAMEWORK

We first give a brief introduction to the quantum probability formalism. We then discuss its application to IR.

3.1 Quantum Probabilities

The quantum probability formalism is a geometric generalisation of standard probability theory that makes use of Hilbert spaces, unit vectors and subspaces. We present the components of the formalism used in this paper.

3.1.1 Systems and States

Quantum theory describes the behaviour of matter at atomic and subatomic scales by means of physical systems which can be in a certain *state* in a *state space*. The latter is a Hilbert space $\mathcal{H}$, a vector space with an inner product. The state of the system is described by a unit vector in the state space called the *state vector*. States determine *statistically* the measures obtained on the system, for instance the position of a particle. Here, the state vector determines the probability that the particle is at a given position.

A system state may be fully known, in which case the system is described by exactly one state vector, but the formalism also allows to be uncertain about its state, in which case the system state can be represented as a weighted set of possible state vectors with corresponding probabilities or, equivalently, as a so-called density operator [16]. Depending on the application we may regard single-part systems, but very often we need to discuss multi-part systems. We therefore introduce single-part systems, including the cases where states are known or uncertain, and then proceed to multi-part ones.

3.1.2 Single-part Systems

States and Probabilities.

Given a state space $\mathcal{H}$ and a state vector φ, a *probabilistic event* is represented as a subspace S of $\mathcal{H}$. A state vector φ induces a probability distribution on events (i.e., the subspaces). The probability of an event is given by the square of the length of the projection of φ onto the corresponding event subspace S, that is by computing the value $\left\|\widehat{S}\varphi\right\|^2$ where $\widehat{S}$ is the projector onto the subspace S. This value is the probability of the event S with respect to the probability distribution defined by φ.

Uncertain States and Weighted Sets.

In a physics experiment, there is often some uncertainty on the preparation process of the system, which in turn induces some uncertainty about the state the system is in at the beginning of the experiment. To formalise this uncertainty, we make use of a *weighted set* of possible state vectors (called ensembles in quantum physics).

A weighted set is defined by a function $\mathcal{V}$ that associates a weight $\mathcal{V}(\varphi) \in \mathbb{R}$ to each of its elements (state vectors) φ. The weight $\mathcal{V}(\varphi)$ corresponds to the probability that the system is in state φ. We say that $\varphi \in \mathcal{V}$ if its weight is greater than zero. For notations, we use standard operations on function spaces: Adding two weighted sets $\mathcal{V}_1 + \mathcal{V}_2$, or scaling by a factor $\alpha\mathcal{V}$. We denote $\varphi \mapsto w$ the weighted set that associates w to φ and 0 to the other vectors.

As states are mutually exclusive (a system cannot be in two different states at the same time), we require that a weighted set representing a system has weights that sum up to 1. Given the weighted set $\mathcal{V}$ of possible state vectors, we can then define the probability of an event S in case the system state is uncertain as

$$\begin{aligned} \Pr(S|\mathcal{V}) &= \sum_{\varphi\in\mathcal{V}} \mathcal{V}(\varphi) \Pr(S|\varphi) = \sum_{\varphi\in\mathcal{V}} \mathcal{V}(\varphi) \left\|\widehat{S}\varphi\right\|^2 \quad (1) \\ &= \operatorname{tr}\left(\rho_{\mathcal{V}}\widehat{S}\right) \text{ where } \rho_{\mathcal{V}} = \sum_{\varphi\in\mathcal{V}} \mathcal{V}(\varphi)\, \varphi\varphi^\top \end{aligned}$$

where tr is the trace operator, $\varphi^\top$ the transpose of φ and $\rho_{\mathcal{V}}$ is the *density operator*, which corresponds to a (probabilistic) *mixture* of states $\varphi \in \mathcal{V}$ (see [16, p. 83ff.]). Equation 1 reduces to $\left\|\widehat{S}\varphi\right\|^2$ if $\mathcal{V}$ consists of only one vector φ, i.e. when there is no uncertainty about the system state.

3.1.3 Multi-part Systems

In physics, interesting systems are those composed of multiple particles. The corresponding systems are *multi-part systems* made of single-part ones. Multi-part systems can be represented in a tensor product of Hilbert spaces, denoted $\otimes$. If $\mathcal{H}_1$ and $\mathcal{H}_2$ are two Hilbert spaces of respective dimensions n and m, the *tensor space* $\mathcal{H}_1 \otimes \mathcal{H}_2$ is an $n \cdot m$-dimensional Hilbert space. If φ_1 is a vector in $\mathcal{H}_1$ and φ_2 is a vector in $\mathcal{H}_2$, then $\varphi_1 \otimes \varphi_2$ is a vector in $\mathcal{H}_1 \otimes \mathcal{H}_2$. Furthermore, if A and B are subspaces (events) in $\mathcal{H}_1$ and $\mathcal{H}_2$, respectively, then $A \otimes B$ is a subspace (event) in $\mathcal{H}_1 \otimes \mathcal{H}_2$. The projection of $\varphi_1 \otimes \varphi_2$ onto $A \otimes B$ is the tensor product of the two projected vectors; if one of the projections is null, then the result is the null vector in $\mathcal{H}_1 \otimes \mathcal{H}_2$. The norm of a vector $\varphi_1 \otimes \varphi_2$ is the product of the norms of the component vectors. From Equation 1, the definition of the projection and the norm in the tensor space allows us to compute the probability of any event. For example, the probability of the composite event $A \otimes B$ is $\Pr(A \otimes B|\varphi_1 \otimes \varphi_2) = \Pr(A|\varphi_1) \times \Pr(B|\varphi_2)$ with φ_1 (φ_2) being a state vector in the first (second) state space.

These operations can be extended to more than two spaces and to weighted sets. In the latter case, with n spaces, the weighted set $\bigotimes \mathcal{V}_i = \mathcal{V}_1 \otimes \ldots \otimes \mathcal{V}_n$ is composed of all tensor combinations of vectors from the weighted sets $\mathcal{V}_i$:

$$\bigotimes \mathcal{V}_i = \sum_{\varphi_1\in\mathcal{V}_1} \cdots \sum_{\varphi_n\in\mathcal{V}_n,} \left(\bigotimes_i \varphi_i \mapsto \mathcal{V}_1(\varphi_1) \times \ldots \times \mathcal{V}_n(\varphi_n) \right) \quad (2)$$

where the weight of an element $\bigotimes_i \varphi_i \in \bigotimes \mathcal{V}_i$ is the product of the weights of its component elements. From Equations 1 and 2, we can compute the probability associated with such a tensor product:

$$\Pr\left(\bigotimes_i S_i \middle| \bigotimes_i \mathcal{V}_i \right) = \prod_i \Pr(S_i|\mathcal{V}_i) \quad (3)$$

An example of the use of tensor products in physics is to describe n particles, for which we measure the probability of each being in a given location in space. The above formula expresses that if the particles are independent from each other, then the probability is the product of the probabilities of the individual particles. Transposed to IR, tensor products are useful to express the constraints (particles) that a relevant document (location in space) should fulfill.

3.2 The Quantum IR (QIR) framework

We now describe our framework for IR, which applies the formalisms known from quantum theory introduced above.

3.2.1 Information Need Spaces

The basic assumption of the QIR framework [12] is that there exists a Hilbert space $\mathcal{H}$ of *pure information needs*, called *information need space*. In this space, a state vector φ or pure information need (IN) reflects a system view of the current IN of the user that completely characterises this user's possible IN.

The concept of a "pure" IN[2] is central to our framework. It can be compared to the notion of "nugget" [5], used in summarisation and question-answering to assess the amount of relevant information a summary or an answer contains.

As elaborated later, documents can be represented as subspaces in the IN space $\mathcal{H}$. An IR system that knows a user's pure IN would be able to determine with certainty how to rank documents, i.e. would know how to compute their probability of relevance. From a geometrical perspective, we posit that a pure IN is *fully* answered by a document if the vector representing the former is contained in the subspace representing the latter. It is *partially* answered if the pure IN vector has a non-null projection onto the document subspace. This is analogous to the view of states and probabilities in a physical system. Consequently, if S_d is a subspace representing document d (and $\widehat{S}_d$ the corresponding projector), then Equation 3 can be applied to estimate the probability of relevance of a document given a pure IN.

Typical to IR is the undeterminism that comes from the fact that the representation of a query is only an approximation of the user's IN, and/or that the query may be ambiguous. This is comparable to the uncertainty about a physical system state. We therefore represent the system's view on the user's IN by a weighted set $\mathcal{V}$, which captures each of the user's possible pure INs. We then compute a ranking of documents by applying Equation 1.

3.2.2 Aspect Spaces and Multi-part Systems

User's INs often consist of several "aspects" that relevant documents should address. For example, in the TREC-8 topic 408, "What tropical storms (hurricanes and typhoons) have caused significant property damage and loss of life?", we can identify two (topical) IN aspects, namely tropical storms and significant damage/loss of life. Each IN aspect can be defined within an IN aspect space, where the state vectors are now called *pure IN aspects*. Examples of pure IN aspects are the vectors representing "hurricane" and "typhoon" for the first IN aspect (tropical storms). We use the terminology *pure IN aspect,* since one pure IN aspect addresses one aspect of the IN (tropical storms) in the same way that a pure IN addresses an IN.

An example of an IN aspect space is the *topical space $\mathcal{T}$,* which in this work is equalled to the standard term space where each term is a dimension. A (simplified) example is

[2]In this paper, we use "pure" information need ("pure IN") to distinguish it from information need ("IN") in its more usual sense in IR.

shown in Figure 1a, where the pure IN aspect "pop music" is represented by the terms "music", "chart" and "hit" of the term space. Note that an IN aspect may be of non-topical nature, but in this paper, since we are experimenting with TREC ad-hoc tasks, we only use a topical space $\mathcal{T}$.

Our basic idea is to regard the whole user's IN as a multi-part system where each component system reflects one aspect of the IN. Each IN aspect is represented in a given Hilbert space $\mathcal{H}_i$ (in this paper, we only consider topical spaces, i.e. $\mathcal{H}_i = \mathcal{T}$). Our IN space is then expressed as a tensor product $\mathcal{H} = \bigotimes \mathcal{H}_i$.

In the IN space $\mathcal{H}$, a document is represented by $S_d^{\otimes} = \bigotimes S_d$ where S_d, whose construction is given in Section 4.1, is the representation of the document in the topical space $\mathcal{T}$. The actual IN is a tensor product of weighted sets $\mathcal{V} = \bigotimes \mathcal{V}_i$ where each $\mathcal{V}_i$ corresponds to one IN aspect of a user's IN in the term space $\mathcal{T}$. We describe several constructions of $\mathcal{V}$ based on a query q in Section 4.2. Given a document representation $S_d^{\otimes}$ and a weighted set $\mathcal{V}$, we then compute the probability of relevance for the document d with Equation 3.

The above abstract construction of a multi-part space comprises an infinite number of IN aspects. However, we assume that all INs can be defined by a finite number of aspects, and we introduce a fake *don't care* pure IN aspect, represented by the state vector[3] $\varphi_\top$. Any document subspace S_d contains the "don't care" pure IN aspect, i.e. $\Pr(S_d|\varphi_\top) = 1$. For example, the pure IN state $\varphi = \varphi_1 \otimes \varphi_2 \otimes \varphi_\top \otimes \varphi_\top \otimes \ldots$ with $\varphi_i \in \mathcal{V}_i$ corresponds to a pure IN composed of two genuine pure IN aspects (φ_1 and φ_2) since $\Pr(\otimes S_d|\varphi) = \Pr(S_d|\varphi_1) \times \Pr(S_d|\varphi_2) \times 1 \times 1 \times \ldots$. We denote $\mathcal{V}_\top = \varphi_\top \mapsto 1$ the IN aspect composed of only the don't care aspect. The don't care aspect can also be used to introduce weights for single aspects, as we will show in Section 4.2.2.

4. REPRESENTING DOCUMENTS AND QUERIES

The previous section introduced the abstract QIR framework. In this section, we make things concrete and describe a methodology to instantiate this framework for the ad-hoc task. We describe how to compute a subspace S_d representing the event that the document d is relevant to a topical aspect (Section 4.1). We then discuss the various possibilities to construct a representation $\mathcal{V}_q$ of a query q. The probability that d is relevant to q is then given by $\Pr\left(S_d^{\otimes}|\mathcal{V}_q\right)$, computed according to Equation 3.

4.1 Representing Documents as Subspaces

We give a quick overview of the construction process. This process was first introduced in [13] in a document filtering scenario. [12] further showed how to construct a document subspace representation by experimenting with a number of strategies and associated parameters. Besides adopting the most promising ones, we also propose new parameters.

The document representation is based on the assumption that a typical document answers various (pure) IN aspects. It also assumes that each document can be split into (possibly overlapping and non-contiguous) semantic fragments, where each fragment addresses an IN aspect. This follows from research in focused retrieval, which states that answers to a query, and hence aspects of it, usually correspond to document fragments (sentences or paragraphs) and not full documents.

As outlined in [12], a document subspace can be created based on the document's fragments, i.e. for each document there is a mapping between them and a weighted set $\mathcal{V}_d$ of pure IN aspects.

S_d is then defined as the subspace spanned by the vectors in $\mathcal{V}_d$. S_d is the smallest subspace such that a document is always a fully relevant answer to an IN aspect it contains, or more formally such that $\Pr(S_d|\varphi) = 1$ for any $\varphi \in \mathcal{V}_d$. The document will be *partially* relevant to a pure IN with a probability that depends on the length of the projection of the pure IN vector onto the subspace S_d, as discussed in Section 3.1.2.

In [12] the weighted set $\mathcal{V}_d$ was constructed using non-overlapping fragments (sentences, paragraphs, section or full document). The best performing approach made use of sentence fragments, but had shortcomings with some collections when sentence detection did not work well. In this paper, we used a methodology where fragments are extracted using a sliding window over the text of the documents, making it independent from error-prone sentence detection algorithms. Denoting $t_1, \ldots, t_M$ the sequence of words of the document, and s the window size, each fragment corresponds to the set of terms $w_{k\times s}, \ldots, w_{k\times(s+1)}$ for $1 \leq k < \frac{M}{s} + 1$.

4.2 Representing Queries as Density Operators

We focus on the construction of a representation for a given query q, which corresponds to calculating a tensor product of weighted sets $\mathcal{V}_q = \otimes \mathcal{V}_i$. We show how to compute this tensor for the queries made of a single term (Section 4.2.1) and queries made of several terms (Section 4.2.2). In contrast to document representation, query construction is a non-trivial task, as there are many ways to compose different terms, each one influencing retrieval performance.

4.2.1 Single-term query

As a query in its simplest form consists of a set of terms, we are first interested in building the query representation for a query composed of a single term, t. This representation is later needed for constructing the representation of multi-term queries.

We assume that a query term t can be represented as the multiset $\mathcal{V}_t$ of pure IN aspect vectors that correspond to document fragments centred on (containing) term t. That is, we use the immediate surroundings of the term t occurrences in documents to build that term representation. This methodology is similar to pseudo-relevance feedback using passages from retrieved documents containing the query terms [1]. The difference is that we use all the passages to build the query representation as we want to consider all possible pure IN aspects associated with the term t. $\mathcal{V}_t$ is then the set of all ambiguous INs represented by t.

As we have *a priori* no way to distinguish between the different vectors in $\mathcal{V}_t$, we assume that each vector is equally likely to be a pure IN aspect with respect to the user's actual IN. Hence, to construct the weighted set $\mathcal{V}_t$ for a term t, for any given pure IN aspect φ, we set

$$\mathcal{V}_t = \sum \varphi \mapsto \frac{n_t(\varphi)}{N_t} \tag{4}$$

[3] $\varphi_\top$ is an extra dimension of each IN aspect space.

where N_t is the number of vectors associated with term t and $n_t(\varphi)$ is the number of occurrences of the pure IN aspect φ in the fragments containing term t. In practice, this representation means that the more vectors φ from $\mathcal{V}_t$ lie in the document subspace, the higher the relevance of the document to the single-term query t.

4.2.2 Multi-term query

We now show three different methods to compute the weighted set corresponding to a multi-term query:

mixture The document should contain as many pure IN aspects of *any* of the query terms;

mixture of superpositions The document should contain as many pure IN aspects which are "combinations" of the pure IN aspects associated with each query term;

tensor product The document should have as many pure IN aspects for *each* of the query terms.

The difference between the mixture and mixture of superpositions approaches and the tensor product one lies in their ability to handle queries containing different aspects. The former two provide no explicit means to distinguish between aspects; they operate in one aspect space and treat each IN associated with a term equally, even if the terms describe different aspects. In contrast, the latter provides explicit means to distinguish between aspects by combining different aspect spaces. As in our test collection we have no indication about which query term address which aspect of the IN, we make the simplifying assumption that each term relates to a different IN aspect.

In the following, we explain each of these three approaches and their rationales. We denote $T = \{t_1, \ldots, t_n\}$ the terms forming the query q, and the indices i and j refer to terms t_i and t_j. As different terms have a different importance, we use a set of weights w_i that sum to 1 to denote the relative importance of the different terms in the query.

Mixture.

We assume that a relevant document should equally answer all pure IN aspects associated with any of the query terms. That is, a document d_1 will have a higher probability of being relevant than d_2, if when we pick by random a term t_i of the query, and then a pure IN aspect φ from the associated set $\mathcal{V}_i$, the probability that d_1 is relevant to the pure aspect φ is in average higher than that of d_2.

More precisely, to compute the probability of relevance of a document d, we first select the i^{th} term of the query (with a probability w_i, used to reflect the importance of the term in the query), and then one of the vectors of the weighted set $\mathcal{V}_i$ corresponding to the term t_i. With this vector, we compute the probability of a document d to be relevant to this aspect. We repeat the process and average over all the possible selections. This defines the probability of relevance of document d given the query. Formally, this corresponds to a density defined as a *mixture* of all the pure IN aspect vectors associated with the query terms. The weighted set is built from the individual weighted set $\mathcal{V}_i$ (Section 4.2.1):

$$\mathcal{V}_T^{(m)} = \sum_{i=1}^{n} w_i \mathcal{V}_i \tag{5}$$

where (m) stands for mixture. In this weighted set, the user's pure IN aspect corresponds to φ with a probability $\sum_i w_i \mathcal{V}_i(\varphi)$. This representation, expressed within a tensor product of IN aspect spaces, makes use of only one of them, and is defined as:

$$\mathcal{V}_q^{(m)} = \mathcal{V}_T^{(m)} \otimes \mathcal{V}_\top \otimes \ldots \tag{6}$$

Note that standard vectorial IR would be derived if $\mathcal{V}_T^{(m)}$ was composed of one pure IN aspect φ and the document subspace was unidimensional.

Mixture of superpositions.

In vector-based IR, a query is represented by a vector that corresponds to a linear combination of the vectors associated with the query terms. In the simplest case, a term vector is naught everywhere but for the component that corresponds to the term, where the value is e.g. tf-idf. In quantum theory, a normalised linear combination corresponds to the principle of superposition (normalised linear combination), where the description of a system state can be *superposed* to describe a new system state.

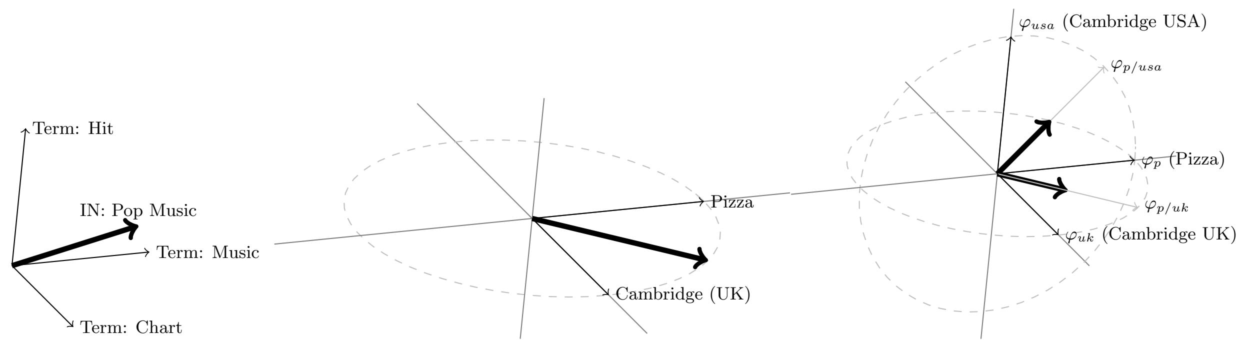
In our case, a system state is a user's pure IN aspect, and we use the superposition principle to build new pure IN aspects from existing ones, as illustrated with the example shown in Figures 1b and 1c. Let φ_p, φ_{uk} and φ_{usa} be three vectors in a three-dimensional IN space which, respectively, represent the pure IN aspects "I want a pizza", "I want it to be delivered in Cambridge (UK)" and "I want it to be delivered in Cambridge (USA)". The pure IN aspect vector "I want a pizza to be delivered in Cambridge (UK)" would be represented by a superposition $\varphi_{p/uk}$ of φ_p and φ_{uk}, as depicted in Figure 1b. We can similarly build the pure IN aspect for Cambridge (USA). To represent the ambiguous query "pizza delivery in Cambridge" where we do not know whether Cambridge is in the USA or the UK, and assuming there is no other source of ambiguity, we would use a mixture of the two possible superposed pure IN aspects, as depicted by the two vectors of the mixture in Figure 1c. This brings us to another variant of query construction, the mixture of superpositions.

To compute the probability of relevance of a document, we randomly select a vector from the set $\mathcal{V}_i$ of each term t_i of the query. In our previous example, the set $\mathcal{V}_1$ would be just one vector (φ_p), whereas $\mathcal{V}_2$ would contain two vectors (φ_{uk} and φ_{usa}). We then superpose the selected vectors (one for each term), where the weight in the linear combination is $\sqrt{w_i}$, obtain a new vector φ, and compute the probability of the document to be relevant to this new pure IN aspect φ.

The above process is repeated for all the possible selections of vectors. The associated weighted set is thus formally defined as the following *mixture of superpositions*:

$$\mathcal{V}_T^{(ms)} = \sum_{\varphi_1 \in \mathcal{V}_1} \cdots \sum_{\varphi_n \in \mathcal{V}_n} \left(\frac{\sum_{i=1}^n \sqrt{w_i}\varphi_i}{\left\| \sum_{i=1}^n \sqrt{w_i}\varphi_i \right\|} \mapsto \prod_i \mathcal{V}_i(\varphi_i) \right) \tag{7}$$

where (ms) stands for "mixture of superpositions". In the weighted set, each pure IN aspect is a linear combination $\sum_{i=1}^n \sqrt{w_i}\varphi_i$ of the IN aspect from each of the terms composing the query which is normalised to ensure it is a unit



(a) A pure IN in a IN/term space (b) A superposition of two pure IN aspects (c) A mixture of two pure IN aspects

Figure 1: Operations in the IN/term space

vector. Each of these linear combinations is associated with a weight which is the product of the weights[4] $\prod_i \mathcal{V}_i(\varphi_i)$.

The mixture of superpositions differs from the mixture in the way we combine the query terms[5]. As an example, consider two terms $T = \{t_1, t_2\}$ that always occur in a document, but half of the time in the same sentences, the other half in distinct sentences. A mixture $\mathcal{V}_T^{(m)}$ retrieves all the documents containing the terms, but a mixture of superpositions $\mathcal{V}_T^{(ms)}$ gives a higher probability to the documents that contain both in the same sentences.

As for mixtures, we only make use of one aspect space within a tensor product of IN spaces, and define the weighted set for mixture of superpositions as:

$$\mathcal{V}_q^{(ms)} = \mathcal{V}_T^{(ms)} \otimes \mathcal{V}_\top \otimes \dots \tag{8}$$

Tensor product of term spaces.

We now suppose that, to be relevant to a query that comprises several aspects of an IN, a document should satisfy ideally all of its aspects. Furthermore, users might give different importance to certain aspects, which motivates the introduction of a weighting scheme for aspects. Both methods discussed above cannot handle aspects. To support aspects explicitely, we discuss a quantum analogue of the "weighted and" (`#wand`) operator proposed in [10]. Aspect spaces and tensor products, introduced in the previous section, are the core components of our approach. Our (simplified) assumption is that the query is made of a set of IN aspects, one for each term of T.

Since we suppose that a relevant document is one that addresses each of the IN aspects associated with the terms of the query, using the tensor product means that the query becomes associated with the weighted set $\mathcal{V}_1 \otimes \dots \otimes \mathcal{V}_n$, where each $\mathcal{V}_i$ now corresponds to the query term t_i. However, this representation gives the same importance to each query term, which led to poor performance (not reported here). This motivates the introduction of a "weighted and". With our notations, this gives a probability of relevance defined by:

$$\Pr(d \text{ is relevant}) = \prod_i \Pr(S_d|\mathcal{V}_i)^{w_i} \tag{9}$$

where the case $w_i = 1$ for any i corresponds to a tensor product $\mathcal{V}_1 \otimes \dots \otimes \mathcal{V}_n$. In general, if all the w_i are integers, then the above equation corresponds to a tensor product as defined by Equation 3, where each $\mathcal{V}_i$ appears w_i times in the tensor product.

However, for a set of arbitrary w_i, the above notation does not correspond to a probability distribution defined on a tensor product. We now present two ways to overcome this. The first one is by transforming Equation 9 so that the new weights are integer values. More precisely, we observe that $\Pr(d \text{ is relevant})^\beta$ does not change the ranking of documents defined by Equation 9 when $\beta > 0$. The value β is chosen so that βw_i can be approximated by an integer value. For example, if a query is composed of two terms t_1 and t_2 with respective weights $w_1 = 0.6$ and $w_2 = 0.4$, then $\beta = 5$, and the query is represented by $\mathcal{V}_1 \otimes \mathcal{V}_1 \otimes \mathcal{V}_1 \otimes \mathcal{V}_2 \otimes \mathcal{V}_2$. The first tensor query representation, referred to as $\mathcal{V}_T^{(T1)}$, can be formally defined as:

$$\mathcal{V}_q^{(T1)} = \bigotimes_i \left(\bigotimes_{j=1}^{\beta w_i} \mathcal{V}_i \right) \otimes \mathcal{V}_\top \otimes \dots \tag{10}$$

The above has the disadvantage that there are βw_i aspect spaces assigned to one single term t_i. This leads us to our second solution, a more elegant one, where a term t_i corresponds to only one aspect. The idea is to use the fake pure IN aspect $\varphi_\top$ defined at the end of Section 3.2.2. To include the state $\varphi_\top$ as one of its possible pure IN aspect states associated with term t_i with a probability $f(w_i)$, we modify $\mathcal{V}_i$ as follows:

$$\mathcal{V}_i^{(T2)} = (1 - f(w_i))\, \mathcal{V}_i + (\varphi_\top \mapsto f(w_i)). \tag{11}$$

The second part of the sum assigns the weight $f(w_i)$ to the don't care pure IN aspect, while the first part scales the previous weights so that the sum of all weights is 1 again. Equations 11 and 1 imply that $\Pr(S_d|\mathcal{V}_i') = f(w_i) + (1 - f(w_i)) \Pr(S_d|\mathcal{V}_i)$. Our second query representation using tensor products is then defined as:

[4]In practice, we used a modified formula for computational reasons (see [12]).

[5]For one-term sets, i.e. $T = \{t\}$, the two representations give the same representation, i.e. $\mathcal{V}_T^{(m)} = \mathcal{V}_T^{(ms)}$.

$$\mathcal{V}_q^{(T2)} = \bigotimes_i \mathcal{V}_i^{(T2)} \otimes \mathcal{V}_\top \otimes \dots \tag{12}$$

We experimented with various heuristically based functions f (not shown here), but none of them gave good results compared to $\mathcal{V}_q^{(T1)}$. This led us to consider a function f that minimises the mean squared error between $\Pr(S_d|\mathcal{V}_i)^w$ and $f(w_i) + (1 - f(w_i)) \Pr(S_d|\mathcal{V}_i)$. It can be shown that[6] the optimal f is defined by $f(w_i) = \frac{3}{2} - \frac{3}{(w_i+1)(w_i+2)}$ for $w_i \in [0, 1]$ (proof omitted). In the next section, we show that both approaches (T1) and (T2) perform similarly, which implies that the latter should be preferred.

5. EXPERIMENTS AND ANALYSIS

The focus of our experiments is to show that the QIR framework achieves performances comparable to that of standard IR models in an ad-hoc setting. We describe our experimental set up, then present our results and their analysis.

5.1 Experimental Setup

Test collections. We used the TREC 1 to 8 collections (with the exception of TREC-4 since it did not contain the "title" field), and the TREC ROBUST 2004 collection.

Queries and weights. All topics were automatically converted into a query using their "title" part. We experiment with the following query representation approaches: **(M)** mixture $\mathcal{V}_q^{(m)}$, **(MS)** mixture of superpositions $\mathcal{V}_q^{(ms)}$, **(T1)** tensor product with $\mathcal{V}_q^{(T1)}$ and **(T2)** with $\mathcal{V}_q^{(T2)}$. To define the weight w_i of query term t_i, following [12], we use normalised IDF.

Document representation. To create the document subspace, we segmented the document using a sliding window approach. Each fragment is processed by stemming and removing stop-words. Finally, to map the resulting sequence of stems to a (unit) vector in the term space, we use the binary weighting scheme, as it performed well in preliminary experiments, and in addition allows faster eigenvalue decomposition computation.

To find an orthonormal basis and projector for the subspace S_d given the multiset of vectors $\mathcal{V}_d$, we performed an eigenvalue decomposition of $\sum_{\varphi \in \mathcal{V}_d} \mathcal{V}_d(\varphi) \varphi\varphi^\top$ as $\sum_{i=1}^D \lambda_i x_i x_i^\top$, where D is the number of eigenvectors with non-null eigenvalues (D is also the dimension of the associated subspace), $\lambda_i > 0$ are the eigenvalues. It can be shown that the vectors x_i form an orthonormal basis of the subspace S_d (proof omitted). Given the latter, the projector $\widehat{S}_d$ onto the subspace S_d associated with document d is then expressed as $\sum_{i=1}^D x_i x_i^\top$. For computational reasons, we bounded D to not exceed a value of 25, which was found in preliminary experiments to perform well.

Term representation. We compute the single-term query density as $\rho_t = \sum_{\varphi \in \mathcal{V}_t} \mathcal{V}_t(\varphi) \varphi\varphi^\top$ where $\mathcal{V}_t$ is the weighted set defined by the ensemble of vectors corresponding to windows of terms of span s centred on each occurrence of term t in the documents. Using eigenvalue decomposition again, we write ρ_t as $\sum_{i=1}^D \lambda_i x_i x_i^\top$. As some terms occur in a great number of documents and the potential rank (dimension of the subspace where ρ_t is defined) of ρ_t can be very high, we approximated it by limiting the number of considered documents to 10,000 and setting a maximum rank of 10 ($D \leq 10$).

As the vectors constructed from the terms occurring in the document sentences are only an approximation of the underlying pure IN vectors, the vectors in $\mathcal{V}_t$ will contain components that should not be assigned to the term t. To reduce this noise, we used a held-out set of documents (20%), denoted $\mathcal{V}_t^*$. For each pure IN aspect φ in $\mathcal{V}_t^*$, we compute the probability $\Pr(S_\varphi|\mathcal{V}_t^*)$ that the document composed of only one sentence represented by φ is relevant with respect to the density ρ_t (that is $\hat{S}_\varphi = \varphi\varphi^\top$). Through an exhaustive search, we selected the dimension K maximising the log likelihood value over all the vectors in $\mathcal{V}_t^*$. In this case, ρ_t is defined[7] as $\sum_{i=1}^K \lambda_i x_i x_i^\top$.

5.2 Results

We report in Table 1 the results, using mean average precision (MAP). We compare the performance of BM25 (with standard parameters, see [15]), TF-IDF (without document normalisation) and, for the QIR framework, those instantiations corresponding to each query construction process, i.e. mixture, mixture of superpositions, and tensor product (T1 and T2). We used a window span of 5, the default setting in [3].

Overall the results were consistent across all collections. The MAP values are below that of BM25 for mixture and mixture of superpositions, and comparable for both tensor approaches. Given the novelty of our framework, and its still unexplored parameters and their effect, we are satisfied with its performance.

The performance of the QIR framework is well above that of a simple TF-IDF model. This shows that, as will be discussed in Section 6, the QIR framework includes some document length normalisation, one based on the IN aspects present in the document. Hence, it does not need to consider explicitly document length.

The approach T1 performed the best, followed by T2, M and MS. The fact that MS works worse than M can be due to the fact that in general terms denote different components of the INs – in [12] it was observed that using mixture of superpositions was better suited for phrase queries. For the tensor approaches, T2 performs similarly to T1 but in three collections it has a lower performance. Given that T2 is more closely related to the QIR framework, the performance of T2 is high enough to consider using it as a basis for interaction.

We analysed the influence of the sliding window span s on the performance of T1, the best performing query representation. In Figure 2, we report the difference in average precision between BM25 (positive values means that T1 was better), for different window spans ranging from 1 to 15. We also report the results for the methodology used in [12, 13], where sentences were used instead of sliding windows. We observe that using a sliding window of 5 gives the best result in terms of both median and variability, hence validating [3].

Finally, we were interested by the influence of the query length (number of terms in T). In Figure 3, we plotted the difference in average precision between BM25 and the performance of the QIR framework, depending on the query length and on the query representation. We kept constant

[6]We assumed a uniform distribution of the probability $\Pr(S_d|\mathcal{V}_i)$

[7]This was not performed for documents, which have a much smaller number of associated vectors. Hence the above process might remove important facets of the document. This was validated experimentally in [12].

	TREC-1	TREC-2	TREC-3	TREC-5	TREC-6	TREC-7	TREC-8	RB-2004
BM25	0.230	0.209	0.282	0.148	0.224	0.182	0.236	0.242
TF-IDF	$0.084^{\dagger}$	$0.041^{\dagger}$	$0.056^{\dagger}$	$0.035^{\dagger}$	$0.088^{\dagger}$	$0.056^{\dagger}$	$0.082^{\dagger}$	$0.074^{\dagger}$
M	$0.205^{\dagger}$	$0.184^{\dagger}$	$0.226^{\dagger}$	$0.115^{\dagger}$	$0.173^{\dagger}$	$0.142^{\dagger}$	$0.165^{\dagger}$	$0.180^{\dagger}$
MS	$0.209^{\dagger}$	$0.167^{\dagger}$	$0.206^{\dagger}$	0.112^{*}	$0.157^{\dagger}$	$0.117^{\dagger}$	$0.159^{\dagger}$	$0.165^{\dagger}$
T1	0.232	$0.195^{\dagger}$	0.281	0.148	0.214	0.182	0.234	0.240
T2	0.222	0.200	$0.259^{\dagger}$	0.139	0.216	0.179	$0.212^{\dagger}$	$0.228^{\dagger}$

Table 1: This table reports mean average precision (MAP). The first line shows the test collection. The second and third lines show the MAP value for BM25 and TF-IDF, respectively. For the query construction, M stands for mixture, MS for mixture of superpositions, T1 and T2 for tensor product. For completeness, significance of the difference with BM25 is shown for the 0.05 level (*) and the 0.01 level (†).

the span of the window (5). We can first observe that in all cases, the performance degrades with longer queries. This shows that we need to improve the representation of multi-term queries. We also note T1 does only degrade slightly with longer queries.

Summarising, using a larger IN space and sliding windows brought dramatic improvement over previous works. Applying tensor products of aspect spaces to explicitely address different aspects of the IN seems to be a good choice; this finding will be subject to further investigation. The results also show that the tensor product representation T2, with a window span of 5, is a valid starting point to exploit further user interactions, although some work is needed on the automatic query representation construction.

6. POTENTIAL OF THE QUANTUM IR FRAMEWORK

In the previous section we showed that our framework can compete with well-established approaches like BM25 in an ad-hoc scenario. This is very promising, and the added complexity of the quantum formalism brings exciting new possibilities to address some of the current IR challenges.

In this section, we discuss the potential of the QIR framework for three IR challenges, namely, interaction, novelty and diversity. We show in this section how those three facets are related in the framework. More precisely, handling novelty and diversity in this framework is a consequence of (1) handling interaction and (2) having a multi-dimensional query and document representation. We discuss (1) and (2), before addressing diversity and novelty.

6.1 Supporting Different Forms of Interaction and Events

If the hypothesis of the existence of an IN space is correct, then any event of interest can be represented as a subspace. This includes the document relevance (used for relevance feedback), and other forms of interactions such as query reformulation or a user click [14].

While interacting with an IR system, users change their point of view, and relevance, contrarily to topicality, is expected to evolve within a search session in two ways [19]: (P1) The IN becomes increasingly specific *from a system point of view*, e.g. when a user types some keywords or clicks on some documents; and (P2) The IN changes *from a user point of view*. The IN becomes more specific as the user reads some documents, or it can slightly drift as user interests do. We discuss how P1 and P2 are supported within our framework by means of projection.

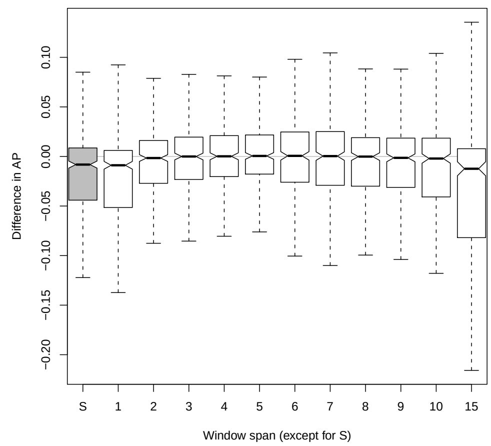


Figure 2: Boxplot of the differences in AP between BM25 and the QIR model (with T1 query representation) with respect to different types of window span and sentence level fragments (S)

Updating Weighted Sets.

We detail how a weighted set, as defined in Section 3.1, is updated when an event is realised. Given the subspace S defining the event, we update the IN given by $\mathcal{V}$ by projecting each of the vectors in $\mathcal{V}$ onto the subspace defined by S, which we write as follows (see [11] for a justification):

$$\mathcal{V} \triangleright S = K \sum_{\varphi / \widehat{S}\varphi \neq 0} \left(\frac{\widehat{S}\varphi}{\left\| \widehat{S}\varphi \right\|} \mapsto \left\| \widehat{S}\varphi \right\|^2 \times \mathcal{V}(\varphi) \right) \quad (13)$$

Here, $\widehat{S}$ is the projector onto the subspace S and K is a normalising factor ensuring that the weights in $\mathcal{V} \triangleright \mathcal{S}$ sum to 1.

The effect of applying Equation 13 on a weighted set $\mathcal{V}$ is that a vector φ orthogonal to S is discarded (the length of the projection is 0), and that the vectors in S are kept as is since $\left\| \widehat{S}\varphi \right\| = \|\varphi\| = 1$. The non orthogonal vectors are projected onto S, and the final weight of each of these vectors depends on how close it was from the subspace S defining the interaction. Geometrically, this means that all the vectors from $\mathcal{V} \triangleright S$ now belong to the subspace S, i.e. the probability of S when the weighted set is $\mathcal{V} \triangleright S$ is 1.

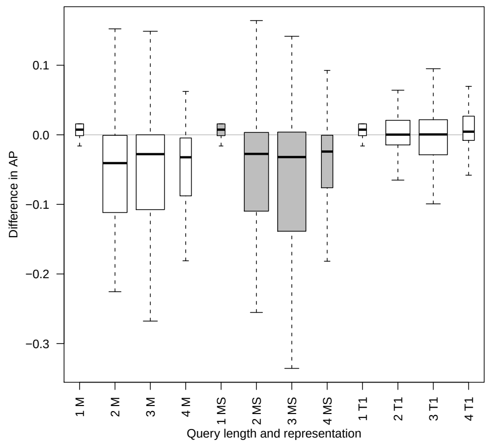


Figure 3: Boxplot of the differences in AP between BM25 and the QIR model (with T1 query representation) with respect to query length (queries with length greater than 4 were grouped with queries of length 4). The width of each boxplot is proportional to the number of topics of a given length.

We illustrate this using Figure 1c, with a weighted set composed of φ_{usa}, $\varphi_{p/usa}$ and $\varphi_{p/uk}$ with respective weights 0.5, 0.2 and 0.3, and considering the subspace S spanned by the vectors φ_p and φ_{uk}. The state φ_{usa} would be discarded, and $\varphi_{p/uk}$ kept as is. Since $\varphi_{p/usa}$ can be written as $\frac{1}{\sqrt{2}}\varphi_p + \frac{1}{\sqrt{2}}\varphi_{usa}$, $\varphi_{p/usa}$ would be projected onto φ_p and $\left\|\widehat{S}\varphi_p\right\| = \left\|\frac{1}{\sqrt{2}}\varphi_p\right\|^2 = \frac{1}{2}$. The new weighted set would be composed of the vectors $\varphi_{p/uk}$ and φ_p with respective weights $0.3K$ and $\frac{1}{2} \times 0.2K$, where K is set to 2.5 so that the sum is 1 (the final weights are respectively 0.75 and 0.25).

Supporting IN Dynamics.

The QIR framework supports both P1 and P2, which we illustrate with the example of Figure 1c. We can define two one-dimensional subspaces $S_{p/uk}$ and $S_{p/usa}$ from the two vectors $\varphi_{p/uk}$ and $\varphi_{p/usa}$. If the user judges a document relevant to "pizzas in the UK", then the weighted set will be projected onto $S_{p/uk}$ and reduced to a weighted set composed of only $\varphi_{p/uk}$. This corresponds to process (P1) where the IN has become more specific. If the user then says that the document about pizzas in US is relevant, the IN will drift towards the pizza-USA direction (it becomes $\varphi_{p/usa}$), hence supporting (P2), and the probability that a document about pizza-UK is relevant will be less than 1 (since the projection of $\varphi_{p/usa}$ onto $S_{p/uk}$ is less than 1).

Event Subspaces.

In practice, to define the subspaces associated with each interaction, we can imagine representations similar to that of queries (Section 4.2). For instance, the query reformulation subspace would be defined by the space spanned by pure IN aspects corresponding to the query terms used in the reformulation, and a click on a document would be represented as a subspace defined by the pure IN aspects corresponding to the terms present in the document snippet.

Negative Feedback.

Even without sophisticated interactions, the QIR framework still allows us to consider the effect of negative relevance feedback, a challenging task in interactive IR. If S_d is the subspace defining the relevance of a document d that is considered to be not relevant by the user, then the event defining the feedback corresponds to the subspace $S_d^{\perp}$ orthogonal to S_d. This is the dual of the approach defined in [17], where a topic is defined by a main vector and a set of vectors corresponding to non relevant documents. In our approach, we start with an initial set of IN vectors and discard those lying in the "not relevant subspace".

6.2 Multi-dimensional Representation of Documents.

We now turn to the multi-dimensional representation of documents. Standard IR models assume that a document deals mainly with one topic. However, a document can cover multiple topics [21], and there is yet no adequate representation of multi-topic documents. Some attempts to integrate topic information have been made in language models (e.g., [18]), but the number of topics is usually small and fixed for a collection, and the topic information is used to smooth a mono-topical language model. In contrast, topics are fined grained (pure IN) in our framework and are not restricted in number – the number of dimensions spanned by a document subspace is a measure of the number of topics discussed in the document.

The mono-topical assumption requires document length normalisation factors like in the BM25 model [15]. We have shown in our experiments that no length normalisation is needed within our QIR. Indeed, when a document covers more topics, it uses more dimensions in the IN space, but if it is centred on a single topic (in the extreme case, only one pure IN), then it should use fewer dimensions. A way to see this is that if we just replace the text of the document by several copies of this text, then the subspace of the document would not change, no matter how many duplicates we create. This means that we capture document length normalisation: the number of dimensions does not change, at least in theory, if the document goes on discussing the same topic.

6.3 Diversity and Novelty

Novelty tackles the task of estimating how novel a document is with respect to previous documents presented to the user. *Diversity* refers to the possible IN aspects associated with a query. With the QIR framework, both novelty and diversity can be dealt with in an unified manner. Indeed, we can express the probability that a new document covers parts of the IN that were not covered by previous ones.

Diversity in Queries.

Diversity in queries is supported because we use weighted sets of INs to represent a query. Indeed, the more diverse this weighted set, the more diverse the query. Recent work using language models [20] has shown that diversity can be improved by considering the different possible "aspects" of a query. Whereas [20] rely on a cost-based formalism, our QIR framework *defines* the query as a set of possible INs.

Handling Novelty and Diversity.

The framework offers a natural way to compute how *relevant and novel* a document d is with respect to a *diverse* user's IN and a set of retrieved documents D. This is done by computing the probability that a document d is relevant to the IN given the user has judged all the documents in D as non relevant. The subspace representing the negation of the latter event is the union $\bigvee_{d' \in D} S_{d'}$ of the subspaces of the documents in D, i.e. the subspaces define the region where answered pure INs do lie. We need to consider the negation of this event, which translates into the orthogonal of the union of the subspaces. We can now define the probability of interest as $\Pr\left(S_d | \mathcal{V} \triangleright \left(\bigvee_{d' \in D} S_{d'}\right)^{\perp}\right)$. In practice, this corresponds to negative relevance feedback (the documents in D are not relevant anymore); thus that strategy cannot be used with standard IR models as they do not handle negative relevance feedback well [17].

7. CONCLUSION

This paper discusses how an interactive probabilistic framework inspired by quantum theory and using the mathematics of Hilbert spaces can be used to tackle contemporary challenges in IR. Documents and information needs are defined within a so-called IN space, which is extended to aspect spaces, enabling us to process different (topical) facets of information needs.

We proposed to define the IN space as a tensor product of smaller term spaces and described a new query representation exploiting this new space. A query is then a conjunction of requirements on these term spaces. We also proposed a new technique to represent documents and terms, based on sliding windows. This technique has the advantage that it does not rely on document markup or sentence detection, in contrast to previous approaches. As a first step, this representation of documents and queries is evaluated in an ad-hoc retrieval scenario, where it is shown that the performance is comparable to standard IR methods like BM25, dramatically improving the performance observed in previous experiments.

Our framework is now mature enough to compete with standard methods in classical IR tasks, but with the prospect of being able to reach far beyond them. The framework includes within the document and information need representation the necessary properties and expressiveness required to handle in a principled way three main IR challenges, namely interaction, diversity and novelty. Exploiting and evaluating further interaction steps, and dealing with novelty and diversity, is part of our future work.

Acknowledgements. This research was supported by an Engineering and Physical Sciences Research Council grant (Grant Number EP/F015984/2). Mounia Lalmas is currently funded by Microsoft Research/Royal Academy of Engineering.

8. REFERENCES

[1] J. Allan. Relevance feedback with too much data. In *SIGIR*. ACM, 1995.

[2] P. Belhumeur, J. Hespanha, and D. Kriegman. Eigenfaces vs. Fisherfaces: recognition using class specific linear projection. *IEEE TPAMI*, 19(7), 1997.

[3] C. Burgess, K. Livesay, and K. Lund. Explorations in context space: Words, sentences, discourse. *Discourse Processes*, 2-3, 1998.

[4] L. Che, J. Zen, and N. Tokud. A ”stereo” document representation for textual information retrieval. *JASIST*, 5, 2006.

[5] C. L. Clarke, M. Kolla, G. V. Cormack, O. Vechtomova, A. Ashkan, S. Büttcher, and I. MacKinnon. Novelty and diversity in information retrieval evaluation. In *SIGIR*. ACM, 2008.

[6] S. Deerwester, S. Dumais, G. Furnas, and T. Landauer. Indexing by latent semantic analysis. *JASIST*, 41(6), 1990.

[7] M. D. Dunlop. The effect of accessing nonmatching documents on relevance feedback. *ACM TOIS*, 15(2), 1997.

[8] P. Ingwersen and K. Järvelin. *The Turn: Integration of Information Seeking and Retrieval in Context (The Information Retrieval Series)*. Springer-Verlag, 2005.

[9] M. Melucci. A basis for information retrieval in context. *ACM TOIS*, 26(3), 2008.

[10] D. Metzler and W. B. Croft. Combining the language model and inference network approaches to retrieval. *IPM*, 40(5), 2004.

[11] M. A. Nielsen and I. L. Chuang. *Quantum computation and quantum information*. Cambridge University Press, 2000.

[12] B. Piwowarski, I. Frommholz, M. Lalmas, and K. van Rijsbergen. Exploring a multidimensional representation of documents and queries. In *RIAO*, 2010.

[13] B. Piwowarski, I. Frommholz, Y. Moshfeghi, M. Lalmas, and K. van Rijsbergen. Filtering documents with subspaces. In *ECIR*, 2010.

[14] B. Piwowarski and M. Lalmas. A quantum-based model for interactive information retrieval. In *ICTIR*, 2009.

[15] S. Robertson and H. Zaragoza. The probabilistic relevance framework: BM25 and beyond. *Foundations and Trends in Information Retrieval*, 3(4), 2009.

[16] C. J. van Rijsbergen. *The Geometry of Information Retrieval*. Cambridge University Press, 2004.

[17] X. Wang, H. Fang, and C. Zhai. A study of methods for negative relevance feedback. In *SIGIR*. ACM, 2008.

[18] X. Wei and W. B. Croft. Lda-based document models for ad-hoc retrieval. In *SIGIR*. ACM, 2006.

[19] H. I. Xie. Shifts of interactive intentions and information-seeking strategies in interactive information retrieval. *JASIS*, 51(9), 2000.

[20] X. Yin, X. Huang, and Z. Li. Promoting Ranking Diversity for Biomedical Information Retrieval Using Wikipedia. In *ECIR*. Springer, 2010.

[21] C. X. Zhai, W. W. Cohen, and J. Lafferty. Beyond independent relevance: methods and evaluation metrics for subtopic retrieval. In *SIGIR*. ACM, 2003.

[22] G. Zuccon and L. Azzopardi. Using the Quantum Probability Ranking Principle to Rank Interdependent Documents. In *ECIR*. Springer, 2010.

[23] G. Zuccon, L. Azzopardi, and C. J. van Rijsbergen. Semantic spaces: Measuring the distance between different subspaces. In *QI*, 2009.