Group-theoretical vector space model

Dohan Kim*

A.I. Research Co., 2537-1 Kyungwon Plaza 201, Sinheung-dong, Sujeong-gu, Seongnam-si, Kyunggi-do, 461-811, South Korea

This paper presents a group-theoretical vector space model (VSM) that extends the VSM with a group action on a vector space of the VSM. We use group and its representation theory to represent a dynamic transformation of information objects, in which each information object is represented by a vector in a vector space of the VSM. Several groups and their matrix representations are employed for representing different kinds of dynamic transformations of information objects used in the VSM. We provide concrete examples of how a dynamic transformation of information objects is performed and discuss algebraic properties involving certain dynamic transformations of information objects used in the VSM.

1. Introduction

Vectors have been widely used in the field of cognitive science [1, 48], machine learning [13], semantics [38], and information retrieval (IR) [26, 52, 54]. The vector space model (VSM) [52, 53] is a model based on a vector space, which represents information objects (e.g., terms, images, documents, queries, etc.) by vectors in a vector space. Each dimension of a vector space represents a feature of an information object, corresponding to a basis element of a vector space of the VSM [38]. A wide variety of weighting schemes [5, 8, 34, 55] have been proposed and tested, in which each component of an information-object vector reflects the importance of the corresponding feature of an information-object vector. For the weighted information-object vectors, distance functions are often used to determine how to measure the similarity between information-object vectors [62]. One common similarity measure between two information-object vectors is the cosine similarity, measuring the cosine of the angle between two information-object vectors in a vector space of the VSM [5]. Besides its intuitive nature, the VSM has also been proven to be effective in IR and relevance ranking [43, 53]. Meanwhile, relevance in IR is often context-dependent as information may evolve with the user, place, and time [43–45], which has not been reflected in the classic, standard VSM [34, 52, 54]. Although the VSM incorporating context and its variants have already been researched [43–45], there is a lack of a systematic approach to representing a dynamic transformation of information objects used in the VSM. Moreover, to the best of our knowledge, theoretical foundations of utilizing group theory for the VSM have not been established. In

---

* Email: dkim@airesearch.kr

this paper we present a group-theoretical VSM and discuss properties on several types of dynamic transformations of information-object vectors in a vector space of the VSM. We also show that some properties are invariant to certain dynamic transformations of information-object vectors. The rest of this paper is organized as follows. Section 2 gives a brief overview of the VSM. In Section 3 we present our group-theoretical VSM to represent a dynamic transformation of information objects used in the VSM. In Section 4 we provide related work and discussion. We give concluding remarks in Section 5. In Appendix we provide the necessary mathematical background on vector spaces, groups, and their representations used in this paper.

2. Vector Space Model (VSM)

In this section we give a brief overview of the classical vector space model (VSM) used in this paper. Vector spaces play an important role in cognitive science [1, 48], semantics [38], pattern classification [13], and information retrieval (IR). In particular, they are commonly used in IR, where IR concerns with methods and procedures of searching and obtaining the required information from information resource or corpus [33, 53]. In IR the *Boolean retrieval model* [40] poses queries having the form of Boolean expression of terms, in which each query consists of terms combined with Boolean operators, such as AND, OR, and NOT. Each document is considered as a set of words in the Boolean retrieval model [40]. However, the Boolean retrieval model has some limitations, such as the lack of similarity measure and a document ranking method [54]. In contrast to the Boolean retrieval model, the VSM has a means to measure the similarity between a query and information-object vectors and to rank information-object vectors according to their similarity scores to the query [40, 54].

The VSM has been formalized as a quadruple $< B, W, S, F >$ [38], where B denotes a set of basis elements of a vector space V of the VSM, W specifies a weight function, S is a similarity measure that maps a pair of information-object vectors to a scalar-valued quantity representing their similarity, and F is a transformation that takes one vector space to another vector space. One of the main purposes of F is to reduce the dimensionality of V [38, 56]. F may also be the identity map that transforms V to itself. Note that a vector space of the VSM is often considered as a *feature space* [58]. Therefore, a wide variety of *feature weighting* (or *scaling*) schemes [37] can be inherited, depending on what kind of a feature space is employed for the VSM.

(1) *Basis Elements B*: B is a set of basis elements $b_1, \ldots, b_n$ that determine the dimensionality of a vector space V of the VSM. Each dimension of V represents a feature of an information object. Each information-object vector v is generated by B, i.e., $v = \sum_{i=1}^{n} w_i b_i$, where w_i's for $i = 1, \ldots, n$ are weights or coefficients. Note that if B' is a set of basis elements $b'_1, \ldots, b'_n$ of V, then v can also be generated by B', i.e., $v = \sum_{i=1}^{n} w'_i b'_i$. One basis can be converted into another basis to reflect a contextual change of information-object vectors, in which a context may refer to the time, space, semantic of information objects, and so on [43, 45]. It means that a basis of a vector space in the VSM can be constructed to represent a context [45].

(2) *Weight function W*: W is a weight function that maps an information object to its normalized form that is often represented as a coordinate vector. Each component of the coordinate vector represents the weight of the corresponding feature of the information object. W is closely related to feature weighting, which relies on the type of a feature space. If a vector space V of the VSM is given as an n-dimensional term space [53], then a query vector Q and a document vector D_i are represented

as $Q = (w_{Q,1}, w_{Q,2}, \ldots, w_{Q,n})$ and $D_i = (w_{i,1}, w_{i,2}, \ldots, w_{i,n})$, respectively. Each term represents each feature of an information object, and each component of an information-object vector represents the importance of a term in a document or query vector [34]. Note that n distinct terms are considered in an n-dimensional term space. There are a wide variety of ways to determine the weight of a term in a given information object. The simplest approach is the *frequency weighting* [37], in which the weight is simply equal to the frequency of a feature. A common approach to term-weighting is the *tf-idf* [34, 40] method, where the weight of a term in a document vector is determined by the local and global factor. The local factor (*term frequency $tf_{i,j}$*) indicates how often term j appears in document i, while the global factor (*inverse document frequency idf_j*) indicates how often term j appears in a document collection [34]. More specifically, the weight of term j in document i for the *tf-idf* method is defined as $w_{i,j} := tf_{i,j} \times idf_j = tf_{i,j} \times \log(N/df_j)$, where N is the total number of documents in a document collection and df_j denotes the number of documents (in a document collection) containing term j [34, 40]. Note that the inverse document frequency ($idf_j := \log(N/df_j)$) assigns a low value to a term that occurs in a large number of documents, while assigning a high value to a term that occurs in a small number of documents in a document collection [34]. The interested reader may also refer to [14, 37, 40] for other term-weighting schemes, such as *Entropy weighting* [37] and *Logarithmic weighting* [14].

(3) *Similarity measure S*: S is a similarity measure that maps each pair of information-object vectors to a scalar-valued similarity score. The angle between a pair of information-object vectors can be used as a simple similarity measure between the pair of information-object vectors. Specifically, the cosine of the angle can be used as a numeric similarity measure (i.e., 1.0 for identical vectors while 0.0 for orthogonal vectors). Furthermore, if two information-object vectors in a vector space $\mathbb{R}^n$ of the VSM are normalized to the unit length, the cosine of the angle between two information-object vectors is simply the inner product of two information-object vectors. Now, the cosine similarity between two information-object vectors v_1 and v_2 is defined as $sim(v_1, v_2) := (v_1 \cdot v_2)/(\|v_1\|\|v_2\|)$, where $v_1 \cdot v_2$ denotes the inner product of information-object vectors v_1 and v_2. Therefore, in terms of the cosine similarity measure, the higher the value of $sim(u_i, u_j)$, the more similar information-object vectors u_i and u_j are. Other methods are also available for the similarity measure based on a distance function. The interested reader may refer to [63] for further details.

(4) *Transformation F*: F is a transformation[1] that transforms one vector space V to another vector space V'.) The main purpose of F is to reduce the dimensionality of V in such a manner that the dimensionality of V' is smaller than the dimensionality of V. The matrix decomposition techniques are often used for dimensionality reduction (i.e., *singular value decomposition* [38] and *QR decomposition* [5]). In some cases it is also possible to reduce the dimensionality in the preprocessing steps (e.g., *stop word elimination* and *stemming* [54]). F can also be the identity transformation that maps a vector space V to itself.

Although the preprocessing and dimensionality reduction steps are often necessary for the VSM, we omit them in this paper. The interested reader may refer to [5, 38, 54] for further details. Unless otherwise stated, B denotes a set of basis elements of a given vector space, W *tf-idf*, S cosine similarity, and F denotes the identity map in $< B, W, S, F >$

---

[1]Since a transformation F is often used for dimensionality reduction, it is distinguished from an (invertible) linear transformation in this paper. Note that an invertible linear transformation (i.e., *isomorphism* [15]) from a vector space V to itself serves as an element of the general linear group $GL(V)$ (see Appendix).

of the VSM used in this paper. We assume that every vector space of the VSM is finite-dimensional in this paper.

Example 2.1. This example illustrates how the *tf-idf* weighting method and the cosine similarity measure of the VSM are applied to document ranking, where each document and a query are represented by *a bag of words* (unordered words with duplicates allowed) [58]. The bag-of-words model is widely used in a document and image representation [19, 20], spam filtering [17], etc. The following figure shows query Q and three documents D_1, D_2, and D_3, each of which is represented by a bag of words.

Q: $\{term_1 \quad term_2\}$
D_1: $\{term_5 \quad term_1 \quad term_1 \quad term_5\}$
D_2: $\{term_2 \quad term_3 \quad term_3 \quad term_6 \quad term_4\}$
D_3: $\{term_2 \quad term_1 \quad term_2\}$

Figure 1. Bag of words for a query Q and documents D_1, D_2, and D_3.

There are six distinct terms in Figure 1. Table 1 shows term weights for each document and query using the *tf-idf* weighting method [26, 40].

Table 1. Term weights for Q, D_1, D_2, and D_3 in Figure 1.

Term weights $w_{i,j} := tf_{i,j} \times idf_j$										
($tf_{i,j}$: term frequency, idf_j : inverse document frequency)										
Total number of documents N=3, $idf_j := \log(N/df_j)$										
(df_j : number of documents containing term j)										
	$tf_{i,j}$							$w_{i,j} = tf_{i,j} \times idf_j$		
Terms	Q	D_1	D_2	D_3	df_j	N/df_j	idf_j	Q	D_1	D_2	D_3
$term_1$	1	2	0	1	2	3/2	0.176	0.176	0.352	0	0.176
$term_2$	1	0	1	2	2	3/2	0.176	0.176	0	0.176	0.352
$term_3$	0	0	2	0	1	3/1	0.477	0	0	0.954	0
$term_4$	0	0	1	0	1	3/1	0.477	0	0	0.477	0
$term_5$	0	2	0	0	1	3/1	0.477	0	0.954	0	0
$term_6$	0	0	1	0	1	3/1	0.477	0	0	0.477	0

Using Table 1, we compute the cosine similarity between Q and D_i for $1 \leq i \leq 3$. Since $\|D_i\| = \sqrt{\sum_j w_{i,j}^2}$ and $\|Q\| = \sqrt{\sum_j w_{Q,j}^2}$, we have

$\|Q\| = \sqrt{0.176^2 + 0.176^2} \approx \sqrt{0.062} \approx 0.249$,
$\|D_1\| = \sqrt{0.352^2 + 0.954^2} \approx \sqrt{1.034} \approx 1.017$,
$\|D_2\| = \sqrt{0.176^2 + 0.954^2 + 0.477^2 + 0.477^2} \approx \sqrt{1.396} \approx 1.182$,
$\|D_3\| = \sqrt{0.176^2 + 0.352^2} \approx \sqrt{0.155} \approx 0.394$.

Since $Q \cdot D_i = \sum_j w_{Q,j} w_{i,j}$, we have

$Q \cdot D_1 = 0.176 \times 0.352 \approx 0.062$,
$Q \cdot D_2 = 0.176 \times 0.176 \approx 0.031$,
$Q \cdot D_3 = (0.176 \times 0.176) + (0.176 \times 0.352) \approx 0.093$.

Now, the cosine similarity measure between query Q and document D_i[1] for $1 \leq i \leq 3$ are computed as follows:

---

[1] By a slight abuse of notation, we use a document (respectively, a query) and its document vector (respectively, query vector) with the same notation in this paper. The distinction is clear from the context.

$$sim(Q, D_1) = (Q \cdot D_1)/(\|Q\|\|D_1\|) \approx 0.062/(0.249 \times 1.017) \approx 0.245,$$
$$sim(Q, D_2) = (Q \cdot D_2)/(\|Q\|\|D_2\|) \approx 0.031/(0.249 \times 1.182) \approx 0.105,$$
$$sim(Q, D_3) = (Q \cdot D_3)/(\|Q\|\|D_3\|) \approx 0.093/(0.249 \times 0.394) \approx 0.949.$$

For the given query Q, D_3 shows the highest rank with $sim(Q, D_3) \approx 0.949$, while D_2 shows the lowest rank with $sim(Q, D_2) \approx 0.105$. Therefore, according to the cosine similarity measure, document D_3 is the most similar to query Q, while document D_2 is the least similar to query Q.

3.   Group Actions on a Vector Space of the VSM

In this section we use several groups to represent dynamic transformations of information objects. We show that some properties are preserved for certain dynamic transformations, in which those dynamic transformations of information objects are represented by a group action on a vector space of the VSM. For the bag-of-words model, we assume that although the content of an information object can be changed by a dynamic transformation, no term can be introduced during a dynamic transformation of information objects. We first describe how an orthogonal group acts on a vector space $V = \mathbb{R}^n$ of the VSM.

For each n, the set of all $n \times n$ orthogonal matrices with real entries forms a subgroup of $GL(n, \mathbb{R})$, denoted by $O(n, \mathbb{R})$, in which a square matrix M is called *orthogonal* if $M^\top = M^{-1}$ [22]. A linear transformation $T : \mathbb{R}^n \to \mathbb{R}^n$ is called an *orthogonal transformation* on $\mathbb{R}^n$ if its transformation matrix in the standard (ordered) basis is an orthogonal matrix with real entries [3]. Orthogonal matrices include rotation and permutation matrices [3].

Proposition 3.1. *Let $V = \mathbb{R}^n$ be a vector space of the VSM. If $O(n, \mathbb{R})$ acts on V by matrix multiplication, then it preserves the cosine similarity between information-object vectors in V.*

Proof. By the definition of the orthogonal group, if $O(n, \mathbb{R})$ acts on V by matrix multiplication, we have $Mv_1 \cdot Mv_2 = (Mv_1)^\top Mv_2 = v_1^\top M^\top Mv_2 = v_1^\top v_2 = v_1 \cdot v_2$ for $M \in O(n, \mathbb{R})$ and $v_1, v_2 \in V$. Since $(Mv)^\top (Mv) = v^\top v$, we have $\|Mv\| = \|v\|$ for $v \in V$. Therefore, an orthogonal matrix preserves both the inner product and the length of information-object vectors. It follows that for information-object vectors $u, v \in V$ and $g \in O(n, \mathbb{R})$, we have $sim(gu, gv) = (gu \cdot gv)/(\|gu\|\|gv\|) = (u \cdot v)/(\|u\|\|v\|) = sim(u, v)$. $\square$

Example 3.1. (*Householder matrix* [6, 29]) Consider a reflection linear operator $R \in GL(V)$ of a vector space $V = \mathbb{R}^n$ of the VSM that reflects each information-object vector through the vector hyperplane that is orthogonal to a unit vector u. The transformation matrix $[R]$ of the linear operator R with respect to the standard (ordered) basis of $\mathbb{R}^n$ is called a *Householder matrix*, and is given by $I - 2uu^\top$. Let H denote $[R]$. Since $HH^\top = I$ (see [6]), we have $H \in O(n, \mathbb{R})$.

For instance, suppose that we have six terms (i.e., $term_1$, $term_2$, $term_3$, $term_4$, $term_5$, and $term_6$) and three documents (i.e., D_1, D_2, and D_3) as shown in Table 1. The 6×3 term-by-document matrix D is denoted as follows:

$$D = \begin{pmatrix} 0.352 & 0 & 0.176 \\ 0 & 0.176 & 0.352 \\ 0 & 0.954 & 0 \\ 0 & 0.477 & 0 \\ 0.954 & 0 & 0 \\ 0 & 0.477 & 0 \end{pmatrix}.$$

Each column corresponds to a document D_j, while each row corresponds to a term $term_i$. Each element d_{ij} in D represents the term weight of $term_i$ associated with document D_j. Let $u = [-\sqrt{2}/2, \sqrt{2}/2, 0, 0, 0, 0]^\top$ and select the vector hyperplane of $\mathbb{R}^6$ that is orthogonal to u. Then, the Householder matrix H' is computed as follows:

$$H' = \begin{pmatrix} 0 & 1 & 0 & 0 & 0 & 0 \\ 1 & 0 & 0 & 0 & 0 & 0 \\ 0 & 0 & 1 & 0 & 0 & 0 \\ 0 & 0 & 0 & 1 & 0 & 0 \\ 0 & 0 & 0 & 0 & 1 & 0 \\ 0 & 0 & 0 & 0 & 0 & 1 \end{pmatrix}.$$

Now, the transformation of D by H' is computed as follows:

$$D' = H'D = \begin{pmatrix} 0 & 0.176 & 0.352 \\ 0.352 & 0 & 0.176 \\ 0 & 0.954 & 0 \\ 0 & 0.477 & 0 \\ 0.954 & 0 & 0 \\ 0 & 0.477 & 0 \end{pmatrix}.$$

The first column of D' represents the transformation of D_1 by H', the second column of D' the transformation of D_2 by H', and the third column of D' represents the transformation of D_3 by H'. It basically replaces $term_1$ with $term_2$, and vice versa[1], in D_1, D_2, and D_3. Since $H' \in O(6, \mathbb{R})$, the cosine similarity among D_1, D_2, and D_3 are preserved among $H'D_1$, $H'D_2$, and $H'D_3$ by Proposition 3.1.

The set of all $n \times n$ invertible diagonal matrices with real entries forms a subgroup of $GL(n, \mathbb{R})$ [49], which is denoted by $D(n, \mathbb{R})$ in this paper. We first describe a *scaling matrix* [7]. A scaling matrix is a diagonal matrix, in which each element in the main diagonal represents a *scaling factor* s_i for the i-th coordinate axis. If $s_i > 1$, it represents a *dilation* transformation in the direction of the i-th coordinate axis. If $0 < s_i < 1$, it represents a *contraction* transformation in the direction of the i-th coordinate axis. If $s_i = -1$, it represents a *reflection* transformation in the direction of the i-th coordinate axis. Note that if a scaling matrix has no zero in its main diagonal, it is invertible.

We say that a linear operator $T : V \to V$ is *diagonalizable scaling linear operator* if there exists an (ordered) basis of V with respect to which the transformation matrix of T is an invertible scaling matrix.

Proposition 3.2. *Let V be an n-dimensional vector space over $\mathbb{R}$ of the VSM. If $D(n, \mathbb{R})$ acts on V by matrix multiplication, $d \in D(n, \mathbb{R})$ represents a transformation matrix of a diagonalizable scaling linear operator of V.*

Proof. By the definition of $D(n, \mathbb{R})$, $d \in D(n, \mathbb{R})$ is a diagonal matrix. Since $D(n, \mathbb{R})$ is a subgroup of $GL(n, \mathbb{R})$, $d \in D(n, \mathbb{R})$ is invertible. It follows that its determinant is not zero. Therefore, each d_i in the main diagonal of d is not zero and may represent a scaling factor in the direction of the i-th coordinate axis. It follows that d is an invertible

---

[1]For a further consideration of a permutation of the basis vectors, consider a symmetric group S_n acting on a vector space $V = \mathbb{R}^n$. Let $B = \{b_1, \ldots, b_n\}$ be a basis of V. Then, S_n acts on V by $g(\sum_i c_i b_i) = \sum_i c_i b_{g(i)}$ for $g \in S_n$, $c_i \in \mathbb{R}$, and $\sum_i c_i b_i \in V$. See [2, 15] for further details.

scaling matrix that represents a transformation matrix of a diagonalizable scaling linear operator of V with respect to a given basis of V. $\qquad\square$

Example 3.2. Each component of an information-object vector can be varied by change of context (e.g., a time-dependent document collection [16, 47]). This example shows the systematic way of changing weights using an invertible scaling matrix. The 6×3 term-by-document matrix D in Example 3.1 was given as:

$$D = \begin{pmatrix} 0.352 & 0 & 0.176 \\ 0 & 0.176 & 0.352 \\ 0 & 0.954 & 0 \\ 0 & 0.477 & 0 \\ 0.954 & 0 & 0 \\ 0 & 0.477 & 0 \end{pmatrix}.$$

Suppose that an invertible scaling matrix S is given below:

$$S = \begin{pmatrix} 2 & 0 & 0 & 0 & 0 & 0 \\ 0 & 3 & 0 & 0 & 0 & 0 \\ 0 & 0 & 2 & 0 & 0 & 0 \\ 0 & 0 & 0 & 1 & 0 & 0 \\ 0 & 0 & 0 & 0 & 1 & 0 \\ 0 & 0 & 0 & 0 & 0 & 1 \end{pmatrix}.$$

Then, the transformation of D by S is computed as follows:

$$D'' = SD = \begin{pmatrix} 0.704 & 0 & 0.352 \\ 0 & 0.528 & 1.056 \\ 0 & 1.908 & 0 \\ 0 & 0.477 & 0 \\ 0.954 & 0 & 0 \\ 0 & 0.477 & 0 \end{pmatrix}.$$

The first column of D'' represents the transformation of D_1 by S, the second column of D'' the transformation of D_2 by S, and the third column of D'' represents the transformation of D_3 by S. By means of the scaling matrix S, the weight of $term_1$ and the weight of $term_3$ in a document collection are multiplied by two, while the weight of $term_2$ in a document collection is multiplied by three. The weight of $term_4$, the weight of $term_5$, and the weight of $term_6$ are invariant under S.

Proposition 3.3. *Let V be an n-dimensional vector space over $\mathbb{R}$ of the VSM. A square matrix $s \in GL(n, \mathbb{R})$ has n linearly independent eigenvectors if and only if it represents a diagonalizable scaling linear operator of V.*

Proof. $(\Rightarrow)$
Assume that a square matrix $s \in GL(n, \mathbb{R})$ has n linearly independent eigenvectors. Since $s \in GL(n, \mathbb{R})$ by assumption, s is invertible. It follows that the determinant of s is not zero. Since s has n linearly independent eigenvectors by assumption and similar matrices have the same determinant, s is diagonalizable to an invertible diagonal matrix $s' \in D(n, \mathbb{R})$ by Theorem A.2[1] and Lemma A.2. Therefore, by Theorem A.3 and Proposition 3.2, it represents a diagonalizable scaling linear operator of V.

---

[1]See Appendix for Theorem A.1–A.4 and Lemma A.1–A.3.

($\Leftarrow$)

If $s \in GL(n, \mathbb{R})$ represents a diagonalizable scaling linear operator of V, then it is diagonalizable by Theorem A.3. Therefore, $s \in GL(n, \mathbb{R})$ has n linearly independent eigenvectors by Theorem A.2. $\qquad\square$

Remarks. The above proof of Proposition 3.3 involves Theorem A.3, which in turn involves a change of basis of a vector space. In [41–45] context is modeled by a basis of a vector space of the VSM. By Proposition 3.3, a certain type of invertible linear operators of a vector space of the VSM can be simplified to a type of diagonalizable linear operators by means of a change of context if context is modeled by a basis of a vector space of the VSM.

Similarly to the above proposition, we have the following lemma by Lemma A.1.

Lemma 3.1. *Let V be an n-dimensional vector space over $\mathbb{R}$ of the VSM. If $s \in GL(n, \mathbb{R})$ is symmetric, it represents a diagonalizable scaling linear operator of V.*

In Example 3.1 we considered a Householder matrix H given by $I - 2uu^\top$, where u is a unit vector orthogonal to the selected vector hyperplane. Since $(I - 2uu^\top)^\top = I^\top - 2(u^\top)^\top(u^\top) = I - 2uu^\top$, it is symmetric. Since the determinant of a householder matrix is -1 [29], we have $H \in GL(n, \mathbb{R})$. Therefore, by Lemma 3.1, H may represent a diagonalizable scaling linear operator of $V = \mathbb{R}^n$.

Example 3.3. Suppose that a feature space $V = \mathbb{R}^4$ of the VSM has four features ($location_1$, $location_2$, $height$, and $brightness$) and some normalized feature vectors. Let $B = \{e_1, e_2, e_3, e_4\}$ denote the standard (ordered) basis of $\mathbb{R}^4$. Now, four features in the feature space are interpreted in such a manner that $e_1 := location_1, e_2 := location_2, e_3 := height$, and $e_4 := brightness$. Suppose also that the transformation matrix $[T]_B$ of a linear operator $T : V \to V$ with respect to B is given as follows.

$$[T]_B = \begin{pmatrix} 3 & 1 & 0 & 0 \\ 1 & 3 & 0 & 0 \\ 0 & 0 & 1 & 0 \\ 0 & 0 & 0 & 1 \end{pmatrix}.$$

Since $[T]_B$ is an invertible and symmetric matrix, it is diagonalizable to an invertible diagonal matrix in $D(4, \mathbb{R})$ by Lemma A.1 and A.2. In other words, there is a transition matrix P from an ordered basis $B' = \{e_1', e_2', e_3', e_4'\}$ to $B = \{e_1, e_2, e_3, e_4\}$ such that $[T]_{B'} = P^{-1}[T]_B P$ is an invertible diagonal matrix, i.e., $[T]_{B'} \in D(4, \mathbb{R})$. By using the diagonalization procedure (see [3]), we have

$$P = \begin{pmatrix} 1/\sqrt{2} & -1/\sqrt{2} & 0 & 0 \\ 1/\sqrt{2} & 1/\sqrt{2} & 0 & 0 \\ 0 & 0 & 1 & 0 \\ 0 & 0 & 0 & 1 \end{pmatrix} \quad \text{and} \quad [T]_{B'} = \begin{pmatrix} 4 & 0 & 0 & 0 \\ 0 & 2 & 0 & 0 \\ 0 & 0 & 1 & 0 \\ 0 & 0 & 0 & 1 \end{pmatrix}.$$

It follows that $e_1' = 1/\sqrt{2}e_1 - 1/\sqrt{2}e_2$, $e_2' = 1/\sqrt{2}e_1 + 1/\sqrt{2}e_2$, $e_3' = e_3$, and $e_4' = e_4$. Therefore, $[T]_{B'} \in D(4, \mathbb{R})$ is a transformation matrix of a diagonalizable scaling linear operator of V with respect to B' by Proposition 3.2.

Let $B(n, \mathbb{R})$ be the set of all $n \times n$ invertible upper triangular matrices with real entries. $B(n, \mathbb{R})$ forms a subgroup of $GL(n, \mathbb{R})$, called the *standard Borel subgroup* [2] of $GL(n, \mathbb{R})$.

Let $V = \mathbb{R}^n$ be a vector space of the VSM and $B = \{e_1, \ldots, e_n\}$ be its fixed standard (ordered) basis. Then, there is an ascending chain of subspaces $\{0\} \subset V_1 = \mathbb{R} \subset V_2 = \mathbb{R}^2 \subset \cdots \subset V_{n-1} = \mathbb{R}^{n-1} \subset V_n = \mathbb{R}^n$, in which each V_i for $1 \leq i \leq n$ is spanned by basis elements $e_1, \ldots, e_i$. This ascending chain is called the *standard complete flag* [2] of $V = \mathbb{R}^n$. The following proposition describes that if information-object vectors in $V = \mathbb{R}^n$ are transformed by an $n \times n$ invertible upper triangular matrix $m \in B(n, \mathbb{R})$, it preserves the standard complete flag of $V = \mathbb{R}^n$.

Proposition 3.4. *Let $V = \mathbb{R}^n$ be a vector space of the VSM and let $\{0\} \subset V_1 = \mathbb{R} \subset V_2 = \mathbb{R}^2 \subset \cdots \subset V_{n-1} = \mathbb{R}^{n-1} \subset V_n = \mathbb{R}^n$ be the standard complete flag of $V = \mathbb{R}^n$. Then, $gV_i = V_i$ for $1 \leq i \leq n$, where $g \in B(n, \mathbb{R})$ is an $n \times n$ invertible upper triangular matrix with real entries.*

Proof. It follows immediately from the fact that the standard Borel subgroup $B(n, \mathbb{R})$ of $GL(n, \mathbb{R})$ stabilizes the standard complete flag of $V = \mathbb{R}^n$ (see [2] for further details). $\square$

Remarks. In IR high-dimensional information-object vectors in a vector space of the VSM are often projected into a low-dimensional subspace in order to improve computational efficiency [58]. Now, consider information-object vectors in a vector space $V_n = \mathbb{R}^n$ of the VSM and project them into a subspace V_i ($1 \leq i \leq n$) of $V_n = \mathbb{R}^n$. By Proposition 3.4, the projected information-object vectors are transformed and remained in that subspace by a linear operator of $V_n = \mathbb{R}^n$ if the transformation matrix of the linear operator with respect to the standard (ordered) basis is an invertible upper triangular matrix with real entries.

For instance, if document D_3 in Example 3.1 is transformed by $t \in B(6, \mathbb{R})$, it still resides in the subspace spanned by the basis element corresponding to $term_1$ and the basis element corresponding to $term_2$. This is not the case if document D_3 is transformed by, let us say, a 6×6 invertible lower triangular matrix formed by replacing the $(3, 1)$-entry of the 6×6 identity matrix with 1.

We next describe the dual space of a vector space V over $\mathbb{R}$ of the VSM. Each information-object vector in V may associate with a scalar-valued quantity. For instance, if a bag of words consists of terms involving product or service items in a *recommender system* [23, 36, 60], each term may associate with a cost (e.g., purchase price). Now, consider the query and documents in Figure 1, where the vector space of the VSM is $V = \mathbb{R}^6$. Let $B = \{u_1, \ldots, u_6\}$ be an ordered basis of $\mathbb{R}^6$. Those six terms are interpreted in such a manner that $u_1 := term_1, \ldots, u_6 := term_6$. Using the frequency weighting scheme, we have $Q = u_1 + u_2$, $D_1 = 2u_1 + 2u_5$, $D_2 = u_2 + 2u_3 + u_4 + u_6$, and $D_3 = u_1 + 2u_2$. We now consider the dual space $\widehat{V}$ of a vector space V of the VSM. Suppose that the costs involving each term are 3 for u_1 and 4, 5, 6, 6 and 7 for u_2, u_3, u_4, u_5, and u_6, respectively. An important linear functional in the dual space $\widehat{V}$ of V is $\phi = 3\hat{u}_1 + 4\hat{u}_2 + 5\hat{u}_3 + 6\hat{u}_4 + 6\hat{u}_5 + 7\hat{u}_6$, in which $< \phi, u_1 > = 3, < \phi, u_2 > = 4$, and so on. By pairing ϕ with a term, the cost of the term is restored. Similarly, by pairing ϕ with an information-object vector, the total cost of the information-object vector is obtained. For instance, the total cost of D_2 is $< \phi, D_2 > = 4 + 2 \times 5 + 6 + 7 = 27$. The following proposition describes the relationship between the representation $\rho : G \to GL(V)$ of G and the dual representation $\hat{\rho}(g) = [\rho(g^{-1})]^\top : \widehat{V} \to \widehat{V}$ of G used in the VSM.

Proposition 3.5. *Let V be a vector space over $\mathbb{R}$ of the VSM and $\widehat{V}$ be the dual space of V. Let $\rho : G \to GL(V)$ be a representation of G and let $\hat{\rho} : G \to GL(\widehat{V})$ be the dual representation of G to $\rho : G \to GL(V)$ acting on $\widehat{V}$ given by*

$$\hat{\rho}(g) = [\rho(g^{-1})]^\top : \widehat{V} \to \widehat{V}.$$

Then, $< \hat{\rho}(g)(\hat{v}), \rho(g)(v) >=< \hat{v}, v >$ for all $g \in G, v \in V$, and $\hat{v} \in \widehat{V}$.

Proof. See Lemma A.3. □

Example 3.4. Let $V = \mathbb{R}^2$ be two dimensional vector space with standard (ordered) basis elements $e_1 = [1, 0]^\top$ and $e_2 = [0, 1]^\top$, and let $\widehat{V}$ be the dual space of V with ordered basis elements $\hat{e}_1$ and $\hat{e}_2$ selected by Theorem A.4. Let u be an information-object vector in V and ψ be a linear functional in $\widehat{V}$ such that $u = e_1 + e_2$ and $\psi = 2\hat{e}_1 + 4\hat{e}_2$. We use the frequency weighting scheme for u and ψ. Therefore, the document corresponding to u consists of the term corresponding to e_1 and the term corresponding to e_2. Similarly, ψ can be interpreted in such a manner that the cost of the term corresponding to e_1 is 2 and the cost of the term corresponding to e_2 is 4. The total cost of u is obtained by pairing ψ with u, that is, $< \psi, u >=< 2\hat{e}_1 + 4\hat{e}_2, e_1 + e_2 >= 6$. Let $\rho : D(2, \mathbb{R}) \to GL(2, \mathbb{R})$ be a matrix representation of $D(2, \mathbb{R})$ associated with V such that $\rho(g) = \begin{pmatrix} 1 & 0 \\ 0 & 2 \end{pmatrix}$ for $g \in D(2, \mathbb{R})$. Then, $\rho(g)$ transforms $u = e_1 + e_2$ into $u' = e_1 + 2e_2$. We see that u' now consists of a single e_1 and two e_2's. Let $\hat{\rho} : D(2, \mathbb{R}) \to GL(2, \mathbb{R})$ be the dual matrix representation of $D(2, \mathbb{R})$ associated with $\widehat{V}$ as shown in Proposition 3.5. We then have $\hat{\rho}(g) = [\rho(g^{-1})]^\top = \begin{pmatrix} 1 & 0 \\ 0 & 1/2 \end{pmatrix}$ that satisfies $< \hat{\rho}(g)(\psi), \rho(g)(u) >=< \psi, u >$. Note that $\rho(g), \hat{\rho}(g) \in D(2, \mathbb{R})$ for $g \in D(2, \mathbb{R})$. Now, $\hat{\rho}(g)$ transforms $\psi = 2\hat{e}_1 + 4\hat{e}_2$ into $\psi' = 2\hat{e}_1 + (4/2)\hat{e}_2 = 2\hat{e}_1 + 2\hat{e}_2$. This means that the cost of the term corresponding to e_2 has to be reduced to the half of the original cost of the term corresponding to e_2 so that the value of $< \psi, u >$ is invariant, i.e., $< \hat{\rho}(g)(\psi), \rho(g)(u) >=< \psi', u' >=< \psi, u >$.

4. Related Work and Discussion

The proper representation of information objects plays an important role in information retrieval (IR), since without it, we cannot expect the good retrieval performance.

This paper has assumed that information objects are represented by vectors in a vector space of the VSM and that certain types of transformations of information objects are well-defined by linear transformations of a vector space of the VSM. The results shown in this paper are concerned with the representation of information objects involving several types of transformations in a vector space of the VSM for the purpose of information retrieval (IR), semantics, etc.

By using group representation theory and linear algebra, this paper provides the mathematical foundation of vector space representation of information objects under group actions, allowing the known group-theoretical results to be adapted for vector space representation of information objects used in IR, semantics, etc.

We have discussed several groups of invertible linear transformations on a vector space of the VSM in previous sections.

In [41–45] context change is modeled by linear transformations from one basis to another in a vector space of the VSM, in order to reflect the information needs evolving with users, time, spaces, etc.

Permutation transformations using permutations of vector coordinates on a *word space* [51] in order to capture and encode word-order information are discussed in [51].

Unitary transformations [28, 59] on a *Hilbert vector space* [31] used in IR are discussed in [59], in which a Hilbert vector space is a *complete inner product space* [31].

In [59] the notions of quantum mechanics (QM) [18], such as *state vector* [18], *observable* [18, 59], *superposition* [4], and *uncertainty* [4, 18], are translated into the notions of IR, intending to apply some of the known theorems (e.g., *Gleason's theorem* [59]) of QM

to the IR context. In that book a document is represented by a vector in a Hilbert vector space, while relevance is represented by a *Hermitian operator* [59] that encapsulates the uncertainty involving relevance. (The interested reader may also refer to [61] for the geometry of conceptual space using vector spaces and quantum theory.)

The dual space model for semantic relations and compositions is discussed in [57], which consists of a domain space and a function space for two distinct similarity measures. However, it does not involve any dual space of linear functionals on a vector space. Meanwhile, the dual space of linear functionals on a vector space is involved in *Dirac notation* [18, 59] that is used for *relevance feedback* [26] and *ostensive retrieval* [59] in IR (see [59]).

Although group representation theory involving *tensor product* [15] of vector spaces are well-studied in mathematics [22], we have not considered any tensor product of vector spaces for semantics in terms of group representation theory in this paper. In computational and mathematical linguistics [10, 57, 58] the vector space tensor product is often used to model *compositionality* [46] (see [24, 25] also for tensor-based compositionality). In [9, 10] a *compositional distributional model of meaning* [9, 10] using *category theory* [39] is discussed, where tensor product is employed for the composition of meanings and types. In that framework VSMs are used for distributional theory of meaning [11], and *Pregroups* [9] are used for a compositional theory for grammatical types [10]. (Both the category of vector spaces and the category of Pregroups are examples of *compact closed categories* [39]. The interested reader may refer to [10] for further details.) We leave it as our future work to consider tensor product representation of certain information objects under group actions for the purpose of IR, semantics, machine learning [13], etc.

5. Conclusions

Although group theory is a major area of research in mathematics, few researches have been done how it is utilized for the VSM. This paper discussed certain dynamic transformations of information objects used in the VSM by means of group-theoretical methods. In our framework an information object is considered as a dynamic entity rather than a static one, where a dynamic transformation of information objects is represented by an element of a group of invertible linear operators on a vector space of the VSM. Several groups act on a vector space V of the VSM by means of their matrix representations, in order to perform a dynamic transformation of information-object vectors systematically. We also showed how the dual space $\widehat{V}$ of V can be employed for the existing VSM. We leave it as an open question to allow other groups that are not discussed in this paper to act on a vector space of the VSM and to derive the useful properties involving some dynamic transformations of information objects used in the VSM.

Appendix. Vector Spaces, Groups, and Representations

In this section we summarize the necessary mathematical background used in this paper. The definitions and results in this section are found in [2, 3, 7, 12, 15, 21, 22, 27, 30–32, 35, 49, 50]. We assume that the reader has some familiarity with linear algebra.

A *group* $(G, \cdot)$ is a nonempty set G, closed under a binary operation $\cdot$, such that the following axioms are satisfied: (i) $(a \cdot b) \cdot c = a \cdot (b \cdot c)$ for all $a, b, c \in G$, (ii) there is a unique element $e \in G$, called the *identity element* of G, such that for all $x \in G$, $e \cdot x = x \cdot e = x$, (iii) for each element $a \in G$, there is an element $a^{-1} \in G$ such that $a \cdot a^{-1} = a^{-1} \cdot a = e$. A group G is *abelian* if its binary operator $\cdot$ is commutative such that $a \cdot b = b \cdot a$ for all

$a, b \in G$.

Let $I_n = \{1, 2, \ldots, n\}$. The group of all bijections $I_n \to I_n$, whose binary operation is function composition, is called the *symmetric group on n letters* and is denoted by S_n.

Let G be a group and H be a nonempty subset of a group G. If H is a group under the restriction to H of the binary operation of G, then H is called a *subgroup* of G.

Let $(G, \cdot)$ and $(G', \circ)$ be groups. A map $\phi : G \to G'$ is a *homomorphism* if $\phi(x \cdot y) = \phi(x) \circ \phi(y)$ for all $x, y \in G$.

A *ring* is a nonempty set R together with two binary operations $+, \times : R \times R \to R$ (called *addition* and *multiplication*) such that: (i) $(R, +)$ is an abelian group, (ii) $(a \times b) \times c = a \times (b \times c)$ for all $a, b, c \in R$, (iii) $a \times (b + c) = (a \times b) + (a \times c)$ and $(a + b) \times c = (a \times c) + (b \times c)$. In addition, (iv) if $a \times b = b \times a$ for all $a, b \in R$, then R is said to be a *commutative ring*, (v) if R contains an element 1_R such that $1_R \times a = a \times 1_R = a$ for all $a \in R$, then R is said to be a ring with *unity*.

If $(R, +, \times)$ is a ring and $(G, \cdot)$ is a group, we also write ab rather than $a \times b$ for $a, b \in R$, and write ab rather than $a \cdot b$ for $a, b \in G$, respectively.

An element x in a ring R with unity is said to be *left* (respectively, *right*) *invertible* if there exists an element z (respectively, $y \in R$) in a ring R such that $zx = 1_R$ (respectively, $xy = 1_R$). An element $x \in R$ that is both left and right invertible is said to be a *unit*.

A ring R with unity $1_R \neq 0$ in which every nonzero element is a unit is called a *division ring*. A *field* is a commutative division ring.

Let R be a ring. A *(left) R-module* is an additive abelian group M together with a *scalar multiplication* defined by a function $R \times M \to M$ such that for all $r, s \in R$ and $a, b \in M$: (i) $(rs)a = r(sa)$, (ii) $(r + s)a = ra + sa$, (iii) $r(a + b) = ra + rb$. In addition, if R is a ring with unity and $1_R a = a$ for all $a \in M$, then M is a unitary *R-module*.

If R is a field, a unitary R-module M is called a *vector space M over R*.

In the remainder of this paper G denotes a group, $\mathbb{K}$ a field, and V denotes a finite-dimensional vector space unless otherwise stated.

Let V, W be vector spaces over $\mathbb{K}$. A function $T : V \to W$ is a *linear transformation* from V to W provided that for all $x, y \in V$ and $k \in \mathbb{K}$: (i) $T(x + y) = T(x) + T(y)$, (ii) $T(kx) = kT(x)$. A linear transformation from V to itself is also called a *linear operator* of V.

A (left) *action* of a group G on a set X is a function $G \times X \to X$ (given by $(g, x) \mapsto gx$) such that for all $x \in X$ and $g_1, g_2 \in G$: (i) $ex = x$, (ii) $(g_1 g_2)x = g_1(g_2 x)$. When such an action is given, we say that G *acts* (left) on the set X.

The *general linear group* $GL(n, \mathbb{K})$ is the group of all invertible $n \times n$ matrices with entries from $\mathbb{K}$ under matrix multiplication. An $n \times n$ matrix is invertible if and only if its determinant is not zero. Alternatively, the general linear group of V is the group of all invertible linear transformations from V to V and is denoted by $GL(V)$. (If V is a finite n-dimensional vector space, then $GL(n, \mathbb{K})$ and $GL(V)$ are isomorphic as groups. See [15] for details.)

The general linear group $GL(n, \mathbb{K})$ and its subgroups act on $V = \mathbb{R}^n$ by matrix multiplication, considering each vector in V as a column matrix. (That is, if $M \in GL(n, \mathbb{K})$ and $x \in V$, $(M, x) \mapsto Mx$.)

A *linear representation* of G is a group homomorphism $\rho : G \to GL(V)$ from G into $GL(V)$. Similarly, a *matrix representation* of G is a group homomorphism $\rho' : G \to GL(n, \mathbb{K})$ from G into $GL(n, \mathbb{K})$.

Suppose G acts on a vector space V over $\mathbb{K}$. The action of G on V is called *linear* if the following conditions are met: (i) $g(v + w) = gv + gw$ for all $g \in G$ and $v, w \in V$, (ii) $g(kv) = k(gv)$ for all $g \in G, k \in \mathbb{K}$, and $v \in V$. If G acts on V linearly, then V itself is called a representation of G, and write gv or $g \cdot v$ for $\rho(g)(v)$.

Let V be a vector space over $\mathbb{R}$. An *inner product* for V is a function $(\ ,\ )$ from

$V \times V$ into $\mathbb{R}$ which satisfies the following for all $x, y, z \in V$ and for all $k \in \mathbb{R}$: (i) $(kx + y, z) = k(x, z) + (y, z)$, (ii) $(x, y) = (y, x)$, (iii) $(x, x) \geq 0$, (iv) if $(x, x) = 0$, then $x = 0$.

Theorem A.1 ([31]). *The equation*

$$(x, y) = \sum_{k=1}^{n} x_k y_k,$$

where $x = (x_1, \ldots, x_n)$, $y = (y_1, \ldots, y_n) \in \mathbb{R}^n$ defines an inner product on $\mathbb{R}^n$.

Let $\|v\| = (\sum_{i=1}^{n} v_i^2)^{\frac{1}{2}}$, where $v = (v_1, \ldots, v_n) \in \mathbb{R}^n$. Then, the geometric interpretation of (u, v) is $(u, v) = \|u\|\|v\|\cos\theta$, where θ is the angle between u and v. For an inner product on $V = \mathbb{R}^n$, we write $u \cdot v$ rather than (u, v)

If we change an ordered basis $B = \{b_1, \ldots, b_n\}$ of an n-dimensional vector space V to the new ordered basis $B' = \{b'_1, \ldots, b'_n\}$, then a vector v has old coordinate matrix $[v]_B$ and a new coordinate matrix $[v]_{B'}$, respectively. It is related to the equation $[v]_B = S[v]_{B'}$, where S is called the *transition matrix* from B' to B. If X and Y are square matrices (i.e., $n \times n$ matrices), then Y is *similar to* X if there is an invertible matrix P such that $Y = P^{-1}XP$.

Let $M = (a_{ij})$ be an $n \times n$ matrix. The *main diagonal* of M consists of the entries a_{ii} for $1 \leq i \leq n$. A matrix D is called *diagonal* if its non-zero entries appear only on the main diagonal. A matrix U is called *upper triangular* if all entries of U lying below the main diagonal are zero. A matrix L is called *lower triangular* if all entries of L lying above the main diagonal are zero.

A square matrix M is called *diagonalizable* if it is similar to a diagonal matrix.

A linear operator T of V is called *diagonalizable* if there exists an (ordered) basis of V with respect to which the transformation matrix of T is a diagonal matrix.

A square matrix is called *symmetric* if $A = A^\top$.

Let V be a vector space over $\mathbb{K}$. If T is a linear operator of V, a nonzero vector $v \in V$ satisfying $Tv = \lambda v$ for some $\lambda \in \mathbb{K}$ is called an *eigenvector* of T. The following theorem describes the fundamental fact of a diagonalizable matrix.

Theorem A.2 ([3]). *An $n \times n$ matrix M with real entries is diagonalizable if and only if M has n linearly independent eigenvectors.*

Lemma A.1 ([3, 7]). *Every symmetric matrix with real entries is diagonalizable.*

Lemma A.2 ([3, 7]). *Similar matrices have the same determinant.*

Given a linear operator $T : V \to V$, the following theorem describes how the transformation matrix of a linear operator of V changes as we change a basis.

Theorem A.3 ([3]). *Let $T : V \to V$ be a linear operator of V and let B and B' be both bases for V. Then, $[T]_B$ and $[T]_{B'}$ are similar, where $[T]_B$ (respectively, $[T]_{B'}$) denotes the transformation matrix of T with respect to B (respectively, B'). Specifically, $[T]_{B'} = S^{-1}[T]_B S$, where S is the transition matrix from B' to B.*

Let V be a vector space over $\mathbb{R}$, and $\mathbb{R}$ be a one-dimensional vector space over itself. Let $\mathrm{Hom}_{\mathbb{R}}(V, \mathbb{R})$ be the set of all linear transformations from V to $\mathbb{R}$. This set, denoted by $\widehat{V}$, forms a vector space over $\mathbb{R}$, which is called the *dual space* of V. Elements of $\widehat{V}$ are called *linear functionals*.

Theorem A.4 ([15]). *If $B = \{v_1, \ldots, v_n\}$ is a basis of a vector space V over $\mathbb{R}$, define $\hat{v}_i \in \widehat{V}$ for each $i \in \{1, \ldots, n\}$ by its action on the basis B in such a manner that $\hat{v}_i(v_j) = \delta_{ij}$ for $1 \leq j \leq n$, where δ_{ij} for $1 \leq j \leq n$ denotes $0 \in \mathbb{R}$ if $i \neq j$ and $1 \in \mathbb{R}$ if*

$i = j$. Then, $\widehat{V}$ is a vector space over $\mathbb{R}$ with basis $\hat{B} = \{\hat{v}_1, \ldots, \hat{v}_n\}$.

There is a (bilinear) natural pairing $<\cdot,\cdot>$ between $\widehat{V}$ and V defined by $<\phi, v> \stackrel{\text{def}}{=} \phi(v)$ for $\phi \in \widehat{V}$ and $v \in V$. (If A denotes a linear operator of V and $A^\top$ denotes its dual or transpose operator of $\widehat{V}$, $(A^\top \phi)(v) = \phi(Av)$ for $\phi \in \widehat{V}, v \in V$ [27].)

Let $\widehat{V} = \mathrm{Hom}_{\mathbb{R}}(V, \mathbb{R})$ be the dual space of V and let $\rho : G \to GL(V)$ be a representation of G. The *dual representation* $\hat{\rho} : G \to GL(\widehat{V})$ to $\rho : G \to GL(V)$ is the representation of G acting on $\widehat{V}$ given by $\hat{\rho}(g) = [\rho(g^{-1})]^\top : \widehat{V} \to \widehat{V}$, where $\hat{\rho}(g)$ is the transpose of $\rho(g^{-1})$.

The following lemma describe the relationship between a representation $\rho : G \to GL(V)$ of G and the dual representation $\hat{\rho} : G \to GL(\widehat{V})$ of G.

Lemma A.3 ([22]). $< \hat{\rho}(g)(\hat{v}), \rho(g)(v) > = < \hat{v}, v >$ *for all* $g \in G, v \in V$, *and* $\hat{v} \in \widehat{V}$.

The definition of the dual representation is such that the following diagram commutes [22]:

$$
\begin{array}{ccc}
V & \stackrel{\phi}{\longrightarrow} & \mathbb{R} \\
{\scriptstyle g}\big\downarrow & & \big\downarrow{\scriptstyle g} \\
V & \stackrel{\longrightarrow}{g\phi} & \mathbb{R}
\end{array}
$$

Therefore, $(g\phi)(v) = g\phi(g^{-1}v)$ for all $g \in G$ and $v \in V$. Since $gx = x$ for all $x \in \mathbb{R}$, we have $(g\phi)(v) = g\phi(g^{-1}v) = \phi(g^{-1}v)$. Since $\phi(g^{-1}v) = ((g^{-1})^\top \phi)(v)$, we have $g\phi = (g^{-1})^\top \phi$, which corresponds to the above definition.

References

[1] J. Aisbett and G. Gibbon, *A general formulation of conceptual spaces as a meso level representation*, Artificial Intelligence 133 (2001), pp. 189–232.

[2] J.L. Alperin and R.B. Bell, *Groups and Representations*, 2nd ed., Springer, New York, NY (1995).

[3] H. Anton, *Elementary Linear Algebra*, 7th ed., John Wiley and Sons, New York, NY (1993).

[4] A. Beiser, *Concepts of Modern Physics*, 5th ed., McGraw-Hill, New York, NY (1995).

[5] M.W. Berry, Z. Drmac, and E.R. Jessup, *Matrices, vector spaces, and information retrieval*, SIAM review 41 (1999), pp. 335–362.

[6] C. Bischof and C. Van Loan, *The WY representation for products of householder matrices*, SIAM Journal on Scientific and Statistical Computing 8 (1987), pp. s2–s13.

[7] O. Bretscher, *Linear Algebra with Applications*, Prentice Hall, Upper Saddle River, NJ (1996).

[8] E. Chisholm and T.G. Kolda, *New term weighting formulas for the vector space method in information retrieval*, Tech. Rep. ORNL/TM-13756, Computer Science and Mathematics Division, Oak Ridge National Laboratory, Oak Ridge, Tennessee, 1999.

[9] S. Clark, B. Coecke, and M. Sadrzadeh, *A Compositional Distributional Model of Meaning*, in *Proceedings of the 2nd Symposium on Quantum Interaction*, March 26–28, Oxford, UK, 2008, pp. 133–140.

[10] B. Coecke, M. Sadrzadeh, and S. Clark, *Mathematical Foundations for a Compositional Distributional Model of Meaning*, in *Lambek Festschrift*, J. van Benthem, M. Moortgat, and W. Buszkowski, eds., vol. 36, Linguistic Analysis, 2010, pp. 345–384.

[11] J.R. Curran, *From Distributional to Semantic Similarity*, Ph.D. thesis, School of Informatics, University of Edinburgh, Edinburgh, Scotland, UK (2004).

[12] K.B. Datta, *Matrix and Linear Algebra*, Prentice-Hall of India, New Delhi, India (1991).

[13] R.O. Duda, P.E. Hart, and D.G. Stork, *Pattern Classification*, 2nd ed., John Wiley and Sons, Hoboken, NJ (2001).

[14] S.T. Dumais, *Improving the retrieval of information from external sources*, Behavior Research Methods, Instruments, & Computers 23 (1991), pp. 229–236.

[15] D. Dummit and R. Foote, *Abstract Algebra*, 3rd ed., John Wiley and Sons, Hoboken, NJ (2004).

[16] J.L. Elsas and S.T. Dumais, *Leveraging temporal dynamics of document content in relevance ranking*, in *Proceedings of the Third ACM International Conference on Web Search and Data Mining*, February 4–6, New York, NY, 2010, pp. 1–10.

[17] M. Erdélyi, A.A. Benczúr, J. Masanés, and D. Siklósi, *Web spam filtering in internet archives*, in *Proceedings of the 5th International Workshop on Adversarial Information Retrieval on the Web*, April 21, Madrid, Spain, 2009, pp. 17–20.

[18] R.P. Feynman, R.B. Leighton, and M. Sands, *Lectures on Physics, Volume III: Quantum Mechanics*, Addison-Wesley, Reading, MA (1965).

[19] D. Filliat, *A visual bag of words method for interactive qualitative localization and mapping*, in *2007 IEEE International Conference on Robotics and Automation*, April 10–14, Roma, Italy, 2007, pp. 3921–3926.

[20] B. Fortuna, D. Mladenić, and M. Grobelnik, *Visualization of text document corpus*, Informatica (2005), pp. 497–502.

[21] J.B. Fraleigh, *A First Course in Abstract Algebra*, Addison-Wesley, Reading, MA (1998).

[22] W. Fulton and J. Harris, *Representation Theory: A First Course*, Springer, New York, NY (1991).

[23] A. Gkatzioura, *Design and implementation of a customer personalised recommender system*, Master's thesis, Universitat Politècnica de Catalunya, Barcelona, Spain (2013).

[24] E. Grefenstette, *Category-Theoretic Quantitative Compositional Distributional Models of Natural Language Semantics*, Ph.D. thesis, Balliol College, University of Oxford, Oxford, UK (2013).

[25] E. Grefenstette and M. Sadrzadeh, *Experimental Support for a Categorical Compositional Distributional Model of Meaning*, in *Proceedings of the 5th international conference on Quantum interaction*, July 27–31, Edinburgh, Scotland, UK, 2011, pp. 1394–1404.

[26] D. Grossman and O. Frieder, *Information Retrieval: Algorithms and Heuristics*, Springer, New York, NY (2004).

[27] B.C. Hall, *Lie Groups, Lie Algebras, and Representations: An Elementary Introduction*, Springer-Verlag, New York, NY (2003).

[28] E. Hoenkamp, *Unitary Operators on the Document Space*, Journal of the American Society for Information Science and Technology 54 (2003), pp. 314–320.

[29] A.S. Householder, *Unitary triangularization of a nonsymmetric matrix*, Journal of the ACM 5 (1958), pp. 339–342.

[30] T. Hungerford, *Algebra*, Springer, New York, NY (1980).

[31] R. Johnsonbaugh and W.E. Pfaffenberger, *Foundations of Mathematical Analysis*, Marcel Dekker, Inc, New York, NY (1981).

[32] D. Kim, *Representations of task assignments in distributed systems using young tableaux and symmetric groups*, arXiv.org arXiv:1012.1288 [cs.DC] (2010), Available at http://arxiv.org/abs/1012.1288v3.

[33] K. Latha and R. Rajaram, *An efficient LSI based information retrieval framework using particle swarm optimization and simulated annealing approach*, in *Proceedings of 16th International Conference on Advanced Computing and Communications*, December 14–17, Chennai, India, 2008, pp. 94–101.

[34] D.L. Lee, H. Chuang, and K. Seamons, *Document ranking and the vector-space model*, IEEE Software 14 (1997), pp. 67–75.

[35] J.M. Lee, *Introduction to Topological Manifolds*, Springer, New York, NY (2000).

[36] G. Linden, B. Smith, and J. York, *Amazon.com Recommendations: Item-to-Item Collaborative Filtering*, IEEE Internet Computing 7 (2003), pp. 76–80.

[37] G. Liu, J. Li, X. Li, and Q. Li, *New Feature Selection and Weighting Methods Based on Category Information*, in *Digital Libraries: International Collaboration and Cross-Fertilization*, Z. Chen, H. Chen, Q. Miao, Y. Fu, E. Fox, and E. Lim, eds., *LNCS*, vol. 3334, Springer Berlin Heidelberg, Shanghai, China, 2004, pp. 330–338.

[38] W. Lowe, *Towards a theory of semantic space*, in *Proceedings of the Twenty-Third Annual Conference of the Cognitive Science Society*, August 1–4, Edinburgh, Scotland, 2001, pp. 576–581.

[39] S. Mac Lane, *Categories for the Working Mathematician*, 2nd ed., Springer, New York, NY (1998).

[40] C.D. Manning, P. Raghavan, and H. Schütze, *Introduction to Information Retrieval*, Cambridge University Press, Cambridge, UK (2008).

[41] R. Mbarek, M. Tmar, and H. Hattab, *A New Relevance Feedback Algorithm Based on Vector Space Basis Change*, in *Computational Linguistics and Intelligent Text Processing - 15th International Conference, CICLing 2014*, April 6–12, Kathmandu, Nepal, 2014, pp. 355–366.

[42] R. Mbarek, M. Tmar, and H. Hattab, *An Optimal Context for Information Retrieval*, in *Algorithmic Aspects in Information and Management - 10th International Conference, AAIM 2014*, July 8–11, Vancouver, Canada, 2014, pp. 323–330.

[43] M. Melucci, *Context modeling and discovery using vector space bases*, in *Proceedings of the 14th ACM international conference on Information and knowledge management*, October 31–November 05, Bremen, Germany, 2005, pp. 808–815.

[44] M. Melucci, *A basis for information retrieval in context*, ACM Transactions on Information Systems (TOIS) 26 (2008).

[45] M. Melucci, *Modeling Retrieval and Navigation in Context*, Information Access through Search Engines and Digital Libraries (2008), pp. 43–57.

[46] J. Mitchell, *Composition in Distributional Models of Semantics*, Ph.D. thesis, School of Informatics, University of Edinburgh, Edinburgh, Scotland, UK (2011).

[47] S. Nunes, C. Ribeiro, and G. David, *Term weighting based on document revision history*, Journal of the American Society for Information Science and Technology 62 (2011), pp. 2471–2478.

[48] E. Roach and B. Lloyd, *Cognition and Categorization*, Lawrence Erlbaum, Hillsdale, NJ (1978).

[49] J.J. Rotman, *An Introduction to the Theory of Groups*, Springer, New York, NY (1994).

[50] B.E. Sagan, *The Symmetric Group: Representations, Combinatorial Algorithms, and Symmetric Functions*, 2nd ed., Springer, New York, NY (2001).

[51] M. Sahlgren, A. Holst, and P. Kanerva, *Permutations as a Means to Encode Order in Word Space*, in *Proceedings of the 30th Annual Meeting of the Cognitive Science Society*, July 23–26, Washington D.C., USA, 2008, pp. 1300–1305.

[52] G. Salton, A. Wong, and C.S. Yang, *A vector space model for automatic indexing*, Communications of the ACM 18 (1975), pp. 613–620.

[53] I.R. Silva, J.N. Souza, and K.S. Santos, *Dependence among terms in vector space model*, in *8th International Database Engineering and Applications Symposium, IDEAS 2004*, July 7–9, Coimbra, Portugal, 2004, pp. 97–102.

[54] A. Singhal, *Modern information retrieval: A brief overview*, IEEE Data Engineering Bulletin 24 (2001), pp. 35–43.

[55] Y.H. Sun, P.L. He, and Z.G. Chen, *An improved term weighting scheme for vector space model*, in *Proceedings of 2004 International Conference on Machine Learning and Cybernetics*, August 26–29, Shanghai, China, 2004, pp. 1692–1695.

[56] R. Tous and J. Delgado, *A Vector Space Model for Semantic Similarity Calculation and OWL Ontology alignment*, in *Database and Expert Systems Applications*, S. Bressan, J. Küng, and R. Wagner, eds., *LNCS*, vol. 4080, Springer Berlin Heidelberg, Kraków, Poland, 2006, pp. 307–316.

[57] P.D. Turney, *Domain and Function: A Dual-Space Model of Semantic Relations and Compositions*, Journal of Articial Intelligence Research 44 (2012), pp. 533–585.

[58] P.D. Turney and P. Pantel, *From frequency to meaning: Vector space models of semantics*, Journal of Artificial Intelligence Research 37 (2010), pp. 141–188.

[59] C.J. van Rijsbergen, *The Geometry of Information Retrieval*, Cambridge University Press, Cambridge, UK (2004).

[60] S. Wang, J. Sun, B.J. Gao, and J. Ma, *Adapting vector space model to ranking-based collaborative filtering*, in *Proceedings of the 21st ACM international conference on Information and knowledge management*, October 29–November 02, Maui, Hawaii, 2012, pp. 1487–1491.

[61] D. Widdows, *Geometry and Meaning*, Center for the Study of Language and Information, Stanford, CA (2004).

[62] S.K.M. Wong and V.V. Raghavan, *Vector space model of information retrieval: a reevaluation*, in *Proceedings of the 7th annual international ACM SIGIR conference on Research and development in information retrieval*, July 2–6, Cambridge, England, 1984, pp. 167–185.

[63] P. Zezula, G. Amato, V. Dohnal, and M. Batko, *Similarity Search: The Metric Space Approach*, Springer, New York, NY (2006).