Topological aspects of information retrieval

EGGHE, Leo & ROUSSEAU, Ronald (1998) Topological aspects of information retrieval. In: Journal of the American Society for Information Science, 49(13). p. 1144-1160.

DOI: 10.1002/(SICI)1097-4571(1998)49:13<1144::AID-ASI2>3.0.CO;2-Z
Handle: http://hdl.handle.net/1942/805

TOPOLOGICAL ASPECTS OF INFORMATION RETRIEVAL

by

Leo Egghe[1] and Ronald Rousseau[2]

ABSTRACT

Let (DS, QS, sim) be a retrieval system consisting of a document space DS, a query space QS, and a function sim, expressing the similarity between a document and a query. Following Everett and Cater, we introduce topologies on the document space. These topologies are generated by the similarity function sim and the query space QS.

Three topolgies will be studied : the retrieval topology, the similarity topology and the (pseudo-)metric one. It is shown that the retrieval topology is the coarsest of the three, while the (pseudo-)metric is the strongest. These three topologies are in general different, reflecting distinct topological aspects of information retrieval. We present necessary and sufficient conditions for these topological to be equal.

[1] LUC, Universitaire Campus, B-3590 Diepenbeek, Belgium and
UIA, Universiteitsplein 1, B-2610 Wilrijk, Belgium
E-mail : legghe@luc.ac.be

[2] UIA, Universiteitsplein 1, B-2610 Wilrijk, Belgium and
KHBO, Zeedijk 101, B-8400 Oostende, Belgium
E-mail : rousseau@kh.khbo.be

Keywords and phrases : topology, information retrieval, similarity, document space, query space.

Severeal examples of topological retrieval systems are presented. One of these examples is a vector space model that yields a simplification of the Everett-Cater model, yet having a more diversified spectrum of topological properties.

Finally, it is shown that information retrieval based on Boolean operators is an intrinsic part of the general topological model. This is a major motivation of the introduction of topologies in theoretical IR-models.

1. INTRODUCTION

Theoretical document retrieval has two main components : indexing and searching. As our article only deals with theoretical aspects we leave all software and hardware aspects aside. Indexing determines the way documents are placed in a database. We will use the term 'indexing' in a very broad way : even when the original document has not changed, e.g., it is directly put on the Internet, we call this a form of indexing. Therefore indexing is considered as a surjective map, f, from the set of original documents, X, to the set of 'indexed document representation', denoted as DS, the document space :

$$f : X \rightarrow DS \qquad (1)$$

Here "surjective" means that every element in DS is the representation of at least one document in X.

Elements of DS will be denoted by D, E, D' etc. They can be bibliographical records with or without abstracts, or full-text documents, or any form of multimedia document.

Searching uses queries and refers to the way customers express an information need, a topic on which they want to be informed. Users' needs are also transformed, and are represented by a formal query. The set of all formal queries is called query space and denoted by QS. So, again we can consider a surjective map, g, from the space of all topics (or use needs) to the query space :

$$g : Y \rightarrow QS \qquad (2)$$

From now on elements in QS will be denoted by the symbols Q, Q' etc... We will mainly be interested in single attribute queries, e.g., keywords or authors. Different query definitions, hence different QS spaces lead to different retrieval systems.

The spaces DS and QS do not completely determine a retrieval system. What is missing is a way to express the degree of agreement between a query $Q \in QS$ and a document $D \in DS$. This is done using a similarity function, denoted as sim :

$$sim : DS \times QS \rightarrow \mathbf{R} : (D,Q) \rightarrow sim(D,Q)$$

We will often keep Q fixed, which yields functions of the form :

$$sim(.,Q) : DS \rightarrow \mathbf{R} : D \rightarrow sim(D,Q) \quad (3)$$

The higher the value of sim(D,Q), the more D and Q correspond and therefore, the more probable it will be that the document D satisfies the users' need. Consequently, such documents are wanted to be in the set of retrieved documents. This retrieved documents set is further determined by a cutoff value $r \in \mathbf{R}$. So, all documents $D \in DS$, such that

$$sim(D,Q) > r \quad (4)$$

are retrieved.

Other forms of thresholds such as

$$r_1 < sim(D,Q) < r_2 \quad (5)$$

or even clustering techniques can also be used (Salton & McGill, 1983).

Everett and Cater (1992) have shown that requirements (4) and (5) can be used to determine topological properties of the space DS. The definition of a topology and of a topological space is given in appendix A. As such it is not clear that such systems express "how near" points (here documents) are to each other. However, many topologies are determined by functions that measure similarity or distance between two documents (see (4), (5) or the example of a metric space as a special case of a topological space - see appendix A) and hence the link with this powerful mathematical tool is clear.

Topologies on DS as determined by (4) or (5), or, in general, by using queries $Q \in QS$, are called topological retrieval systems. Indeed, the above described available notion of "how near" documents are, can be used in IR. It is a kind of predetermined system of documents for which the notion of "relatedness" is available and this independent from a used query Q (i.e. prior to the retrieval action itself, but with the knowledge of QS, the set of all possible queries).

Different topologies will, in addition, yield different retrieval properties. As such the notion of a topological retrieval system is closely linked to the dynamics of retrieval behavior. The mathematical tools made available through these systems allow us to give answers to questions such as: "How do slight alterations in queries or thresholds influence the retrieval result.

In the next section we will set up a general framework for topological retrieval. Three topologies on DS will play a role : the retrieval topology τ, the (pseudo-)metric topology τ' (both introduced by Everett and Cater (1992)), and the similarity topology τ''.

In the second section we present examples of document spaces DS for which the three topologies coincide, as well as examples where any two of these topologies are different. These properties illustrate the specific role each of these topologies plays. In particular, we give an example of a vector space model on DS = I^n (I denotes the unit interval [0,1], $n \in N$) with the inner product as similarity function. We will show that this approach leads to a simpler model, with more interesting propeties, than the vector model studied in (Everett & Cater, 1992) which is based on Salton's cosine measure.

The last section shows that information retrieval based on Boolean connectives AND and OR is an intrinsic part of the general topological model. In particular it is only necessary to have a query space QS consisting of elementary queries. The Boolean aspects are taken care of by the topology of DS. This generalizes the Boolean algebra results of Cater (1986) and (Everett & Cater, 1992).

Mathematical proofs of most of the results are relegated to the appendices.

Notation : We will use the abbreviation iff for the phrase 'if and only if'.

2. DOCUMENT SPACES AS TOPOLOGICAL SPACES

A retrieval system is a triple (DS, QS, sim) consisting of a document space DS, a query space QS and a similarity function sim. At this moment DS is only a set. However, using the query set QS and the similarity function sim will enable us to put an additional structure on DS, namely that of a topological space. We focus here on the document space as a topological space (by using QS) since we want to have a notion of similarity between documents. But theoretically, there is no real difference between QS and DS: we could study a topological system on QS, determined by DS. This is an example of a dual situation. For more on this, see Egghe and Rousseau (1997 a).

At this point we refer the reader who is not familiar with topological notions again appendix A and the second section of (Everett & Cater, 1992). Information on general topology can be found in any of the following books : (Császár, 1978), (Dugundji, 1966) and (Willard, 1970). Special examples of topological spaces can be found in (Steen & Seebach, 1978). Finally, (Wilansky, 1970) and (Kreyszig, 1978) are good references on normed spaces.

In this article we will study three topologies on DS.

The retrieval topology τ (Everett & Cater, 1992)

The retrieval topology, denoted as τ, is generated by the subbasis

$$\{R(Q,r) \mid r \in \mathbf{R}, Q \in QS\} \tag{6}$$

where a retrieval R(Q,r) is defined as :

$$R(Q,r) = \{D \in DS \mid sim(D,Q) > r\} \quad . \tag{7}$$

A subbasis is a subset C of the set of all open sets such that all open sets can be constructed by forming finite intersections of elements of C.

Note that, usually, the range of all sim(.,Q) is in $\mathbf{R}^+$ (the positive real numbers, including zero), so that in these cases τ is generated by

$$\{R(Q,r) \mid r \in \mathbf{R}^+, Q \in QS\}$$

There is no mathematical reason to limit the range of sim to the positive numbers. Therefore, we decided to use $\mathbf{R}$ in theoretical arguments, but to use $\mathbf{R}^+$ in all examples. The set

$$ret_\tau(Q) = \{R(Q,r) \mid r \in \mathbf{R}\} \tag{8}$$

is called the set of retrievals of Q (or, in short, the retrievals of Q) w.r.t. τ. Also the sets R(Q,r), $r \in \mathbf{R}$, $Q \in$ QS are called retrievals of the system (DS, QS, sim, τ) or, in short, of τ.

Note that the set of retrievals of a query Q is the set of all possible answers to the query Q in the system (DS, QS, sim) with the retrieval topology. Different answers are obtained by changing the threshold. Consequently, the retrievals of a fixed query Q, form a nested set, i.e.

$$R(Q,r_1) \subset R(Q,r_2) \quad \text{iff} \quad r_1 \geq r_2$$

Another name for τ could be the "threshold" topology, a name that is clear from the definition (7). Here a document is retrieved if its similarity with a query Q is at least a a certain value r.

The (pseudo-)metric topology τ' (Everett & Cater, 1992)

Everett and Cater (1992) give the following definition. Let (DS, QS, sim) be a retrieval system. Define a function d on DS × DS as follows :

$$d(D,E) = \sup_{Q \in QS} |sim(D,Q) - sim(E,Q)| \tag{9}$$

The function d should be a pseudo-metric (see appendix A for the definition), leading to the pseudo-metric topology τ'. However, there is a problem with definition (9). It is possible that the supremum is infinite, in which case d does not exist in the usual sense. To correct this we will define τ' in a more accurate way. For the ease of the notation, we first define

$$\rho_Q (D,E) = |\text{sim } (D,Q) - \text{sim } (E,Q)| \tag{10}$$

here $D,E \in DS$, $Q \in QS$. Then define

(i) if the supremum of (10) is finite :

$$d (D,E) = \sup_{Q \in QS} \rho_Q (D,E)$$

i.e. the pseudo-metric defined by Everett and Cater;

(ii) $d' : DS \times DS \to [0,1]$ by

$$d'(D,E) = \sup_{Q \in QS} \frac{\rho_Q (D,E)}{1 + \rho_Q (D,E)} \tag{11}$$

(iii) $d'' : DS \times DS \to [0,1]$ by

$$d''(D,E) = \sup_{Q \in QS} [\min (1, \rho_Q (D,E))] \tag{12}$$

Clearly d' and d" always exist and are finite. The following theorem states that from a topological point of view the pseudo-metrics d, d' and d" are 'the same' (although they are, of course, different as metrics).

Theorem 2.1 :

Let d, d' and d" be as defined in (9), (11) and (12). Then

(i) the (pseudo-)metrics d' and d" are equivalent

(ii) if d exists, it is equivalent with d' and d".

The proof of this theorem is given in Appendix B. The topology generated by d (if it exists), d' or d" is called the pseudo-metric topology on DS and is denoted by τ'.

The topology τ' measures absolute distances between documents, independent of a used query in IR.

As noted by Everett and Cater (1992) the following result is true :

Proposition 2.2 :

τ' is generated by the subbasis

$$V(D,\varepsilon) = \{E \in DS | \rho_Q(D,E) < \varepsilon, \forall Q \in QS\}$$

where ε is any strictly positive real number.

The proof follows readily from the proof of Theorem 2.1.

Definition 2.3 (Everett & Cater, 1992) :

The retrieval system (DS, QS, sim) separates the points of DS if,

$$(\forall Q \in QS : sim(D,Q) = sim(E,Q)) \Rightarrow (D = E)$$

It is obvious that the pseudo-metrics above are metrics iff the retrieval system (DS, QS, sim) separates the points of DS.

The similarity topology τ"

In order to obtain consistent results, stability is very important in IR. It is, e.g., necessary that if documents D and E are 'alike' (an expression to be defined in each concrete case) their similarity values must also be close to each other. In this way slight changes in e.g. recall requirements yield only slight changes in the retrieval result. Such a behavior can only be guaranteed if the functions sim(.,Q) are continuous. We therefore define the topology τ", referred to as the similarity topology, as the coarsest topology on DS that makes all similarity functions sim(.,Q) continuous. The next theorem describes a subbasis of the similarity topology.

Theorem 2.4 :

The similarity topology τ'' is generated by the subbasis

$$\{U(Q, r_1, r_2) \mid Q \in QS,\ r_1 < r_2\}$$

where

$$U(Q, r_1, r_2) = \{D \in DS \mid r_1 < sim(D, Q) < r_2\} = sim(., Q)^{-1}(]r_1, r_2[)$$

with $r_1 < r_2$, $r_1, r_2 \in \mathbf{R}$, $Q \in QS$.

Proof :

The proof is easy : it follows immediately from properties of the inverse relation and the fact that the open intervals $]r_1, r_2[$ are a basis for the Euclidean topology on the real line. □

The set

$$ret_{\tau''}(Q) = \{U(Q, r_1, r_2) \mid r_1 < r_2\}$$

is called the set of retrievals of Q (or, in short, the retrievals of Q) w.r.t. τ''. Also the sets $U(Q, r_1, r_2)$, $r_1 < r_2$, $r_1, r_2 \in \mathbf{R}$, $Q \in QS$, are called retrievals of the system (DS, QS, sim, τ'') or, in short, of τ''.

Note that if $r_3 < r_1 < r_2 < r_4$ then

$$U(Q, r_1, r_2) \subset U(Q, r_3, r_4)$$

The topology τ'' describes the retrieval of documents according to their closeness to a given query Q. Here (in contrast with τ) the exact query Q must be matched.
When it is clear from the context with which topology we work we will drop the subscript τ or τ'' and simply write ret(Q).

Results on these topologies will follow in the next sections. We can, however, already point out one intrinsic property of these topological spaces. The topologies τ, τ' and τ'' determine neighborhoods around every document $D \in DS$. These neighborhoods determine a filtering of documents : the finer we look, the more closely we end up in the neighborhood of D. The availability of such neighborhoods determine the degree of "fine tuning" that is possible in IR-systems that work with τ, τ' or τ''. Otherwise stated, the topologies on the document space DS determine a pre-defined structure on DS of "what documents are close to (a) certain document (s)", even without the formulation of a specific query $Q \in QS$. It is the totality QS of all possible queries that determines this pre-defined structure. This in turn is comparable (but totally different in nature) with the statistical clustering techniques for documents that exist - see e.g. Salton and McGill (1983).

The three topologies we have introduced are related in the following way.

Theorem 2.5 :

$$\tau \subset \tau'' \subset \tau'$$

Proof :

The sets $R(Q,r) = sim(.,Q)^{-1}(]r,+\infty[)$, generating the retrieval topology, are open in τ'' since sim(.,Q) is continuous on (DS,τ''). Hence $\tau \subset \tau''$. Furthermore, it is clear that sim(.,Q) is continuous on (DS,τ'), hence $\tau'' \subset \tau'$ as τ'' is the coarsest topology that makes all sim(.,Q) continuous. □

It will be shown in this and in the next section that it is possible that $\tau \neq \tau''$, $\tau'' \neq \tau'$ and even $\tau \neq \tau'' \neq \tau'$ are possible, as well as $\tau = \tau'' = \tau'$!

Everett and Cater (1992) claimed the following results :

Statement 1.

For any $Q \in QS$ and $t \in [0,1]$, the set $U(Q,t) = \{D \in DS \mid sim(D,Q) < t\} \in \tau$.

Statement 2.

Let τ_1 and τ_2 be the retrieval topologies of two essentially equivalent retrieval systems. Assume that (DS, τ_2) is compact, then $\tau_1 = \tau_2$.

Recall (Everett & Cater, 1992) that two retrieval systems (DS, QS, sim_1) and (DS, QS, sim_2) are said to be essentially equivalent if $sim_1(D,Q) < sim_1(E,Q)$ iff $sim_2(D,Q) < sim_2(E,Q)$. Essentially equivalent models retrieve all documents in the same order.

In (Egghe & Rousseau, 1997 b) we have shown that these statements (Lemma 1 and Theorem 4 in (Everett & Cater, 1992)) are wrong. We briefly repeat the counterexamples constructed in (Egghe & Rousseau, 1997 b) and add some new comments.

A counterexample to statement 1.

Let DS = **N** (all natural numbers, including zero); QS can be any non-empty set. Let sim_1 be defined as follows : $sim_1(D,Q) = 1/5$ for all $Q \in QS$, except for one special query Q_1. For this special query $sim_1(n,Q_1)$, $n \in \mathbf{N}$, is given by the following table :

n	$sim_1(n,Q_1)$
0	1/5
1	1/4
3	3/8
5	7/16
7	15/32
and so on for the odd numbers	
2	1/2
4	3/4
6	7/8
and so on for the even numbers	

Note that the similarity values for the odd numbers converge to 1/2; the similarity values for the even numbers converge to 1.

Now the set $U(Q_1,1/4) = \{D \in DS \mid sim_1(D,Q_1) < 1/4\} = \{0\}$ does not belong to the retrieval topology τ_1. Indeed, the sets of τ_1 are the following :

N and ϕ (the empty set),

all natural numbers except zero,

all natural numbers except zero and the j smallest odd numbers (j = 1,2,...)

all natural numbers except zero, the odd numbers and the j smallest even numbers (j = 1,2,3,...).

Note that the set consisting of all even numbers {2,4,6,...} is NOT an open set for this retrieval topology.

This example also shows that the function $sim_1(.,Q_1)$ is not continuous for the retrieval topology. Hence this is also a case where the retrieval topology differs from the similarity topology.

Counterexample to Statement 2.

We consider the same retrieval system as before but - for the second similarity function - make a slight change to the first one. The similarity function sim_2 is everywhere equal to sim_1, except for the value in $(2,Q_1)$. We set $sim_2(2,Q_1)$ equal to 5/8. It is now clear that the models (DS = **N**, QS, sim_1) and (DS = **N**, QS, sim_2) are essentially **equivalent**. Moreover, as $sim_i(0,Q)$ is 1/5 for every Q, i = 1, 2, the set DS = **N** is compact for the retrieval topology (the point zero plays the role of D_0 in the Proposition of (Egghe & Rousseau, 1997b). However, the two retrieval topologies do not coincide. For τ_2 (the retrieval topology derived from sim_2), the set consisting of all even numbers = $\{D \in DS \mid sim_2(D,Q_1) > 1/2\}$ is clearly open (an element of τ_2). We noted before that this set is not open in τ_1. Hence the two topologies are not equal, which contradicts statement 2.

We consider now he following question : can statements 1 and 2 be adapted so that they become true? The answer to this is yes, essentially by using τ" instead of τ. This will be shown in Theorems 2.8.1 and 2.8.2. We first state a simple lemma and introduce a new notion.

Lemma 2.6 :

For any $Q \in$ QS and any $r \in$ **R**, the set

$\{D \in DS \mid sim\ (D,Q) < r\}$

belongs to the similarity topology τ".

Proof :

This follows readily from the definition of τ" and the fact that the sets $]-\infty, r[$ are open in **R**, for every $r \in$ **R**. □

Lemma 2.6 shows that Lemma 1 in (Everett & Cater, 1992) is true for τ". Also their Theorem 4 becomes true when using τ" instead of τ - see theorem 2.8.2 further on.

Definition 2.7 :

We say that a retrieval system (DS, QS, sim) satisfies the maximum principle if $\forall r \geq 0$, $\forall Q \in$ QS, the set

$$A = \{sim(D,Q) \mid D \in DS,\ sim(D,Q) \leq r\}$$

has a maximum, i.e., there exists $D_0 \in$ DS such that $sim(D_0,Q) = \max A$. An analogous definition can be given for the minimum principle.

Examples.

We formulate some cases of retrieval systems satisfying Definition 2.7 (min as well as max principle).

(i) If the range of sim(.,Q) is finite for every $Q \in$ QS then the system satisfies the minimum as well as the maximum principle. This is the case for the classical example where sim can only take values in $\{0,1\}$. This is also the case for a finite document space.

(ii) If DS is compact for a certain topology S such that all sim(.,Q) are continuous (e.g. τ") then the system also satisfies the minimum and the maximum principle. Indeed, under these conditions sim(DS,Q) is compact in $\mathbf{R}$ for every $Q \in$ QS. Consequently sim(DS,Q) is closed and bounded and therefore has a minimum and a maximum.

This brings us to Theorems 2.8.1 and 2.8.2.

Theorem 2.8.1 :

If (DS, QS, sim_1) and (DS, QS, sim_2) are essentially equivalent retrieval systems and if one of the two satisfies the maximum principle then their retrieval topologies are equal.

Proof :

If one model satisfies the maximum principle then, by the fact that they are essentially equivalent, the other model does too (and the maxima are reached for the same document D_0). For every $Q \in$ QS and $r \in \mathbf{R}$, let $D_0 \in$ DS be such that $sim_1(D_0,Q) = \max\{sim_1(E,Q); sim_1(E,Q) \leq r\}$. Then the following expressions are equivalent :

$$D \in R_1(Q,r) = \{E \in DS \mid sim_1(E,Q) > r\} \quad \Leftrightarrow \quad sim_1(D,Q) > sim_1(D_0,Q)$$
$$\Leftrightarrow \quad sim_2(D,Q) > sim_2(D_0,Q)$$
$$\Leftrightarrow \quad D \in R_2(Q,sim_2(D_0,Q))$$

The same argument can be given when the indices are interchanged. Since the sets $\{R_1(Q,r) \mid Q \in QS, r \in \mathbf{R}\}$, $i = 1,2$ form a subbasis for the respective retrieval topologies, these topologies are the same. □

Theorem 2.8.2 :

If (DS, QS, sim_1) and (DS, QS, sim_2) are essentially equivalent retrieval systems and if one of the two satisfies the minimum principle and if one of the two satisfies the maximum principle then their similarity topologies are equal.

Proof :

Since the retrieval systems are essentially equivalent, they both satisfy the minimum and the maximum principle. In the same way as in the proof of theorem 2.8.1 we can now show that, for every $Q \in$ QS and $r_1, r_2 \in \mathbf{R}$, $r_1 < r_2$:

$$\{E \in DS \mid r_1 < sim_1(E,Q) < r_2\}$$
$$= \{E \in DS \mid sim_2(D_1,Q) < sim_2(E,Q) < sim_2(D_0,Q)\}$$

where D_0 resp. D_1 are the documents featuring in the definition of the minimum resp. maximum principle of the first system. Hence

$$U_1 (Q, r_1, r_2) = U_2 (Q, sim_2 (D_1, Q), sim_2 (D_0, Q))$$

Again, the same argument can be given with the indices interchanged and hence $\tau_1'' = \tau_2''$. □

The topology τ'' is defined as the coarsest topology on DS that makes all the functions sim(.,Q) : DS → **R** continuous. Here, **R** has the usual Euclidean topology $\mathscr{E}$. Our next result shows that the retrieval topology can be defined in a similar way, but using a different topology on **R**.

Theorem 2.9 :

Equip the real line **R** with the topology, S, generated by the open half-lines $]r, +\infty[$, $r \in \mathbf{R}$. Then the retrieval topology τ is the coarsest topology on DS that makes all functions

$$sim (., Q) : DS \rightarrow \mathbf{R}, S$$

continuous.

Proof :

The functions sim(.,Q) are continuous from DS to **R**,S iff the sets

$R(Q,r) = sim(.,Q)^{-1}(]r, +\infty [)$ are open in DS. □

One could also say that on (DS, τ) all sim's are lower semi-continuous into **R**, equiped with $\mathscr{E}$ (see Willard (1970), p.49, 7K).

Also the (pseudo-)metric topology τ' can be characterized through continuity properties of the similarity functions.

Theorem 2.10 :

The following assertions are equivalent :

(i) $\tau' = \tau''$

(ii) The family {sim(.,Q) | Q ∈ QS} is equicontinuous on (DS, τ''), where the functions range in $\mathbf{R}, \mathscr{E}$.

For the proof we refer to Appendix C.

Corollary 2.11 :

If QS is finite then $\tau' = \tau''$.

Based on the above theorem and/or Theorem 2.5 we obtain the following easy results.

Proposition 2.12 :

The following assertions are equivalent :

(i) $\tau = \tau''$

(ii) All functions sim(.,Q) are continuous on (DS, τ), with range in $\mathbf{R}, \mathscr{E}$.

Proposition 2.13 :

The following assertions are equivalent :

(i) $\tau = \tau' = \tau''$

(ii) The family {sim(.,Q); Q ∈ QS} is equicontinuous on (DS, τ) with range in $\mathbf{R}, \mathscr{E}$.

3. EXAMPLES OF TOPOLOGIES ON DOCUMENT SPACES

3.1. The counterexample to Statement 1

We repeat that the counterexample to Statement 1 is an example for which $\tau = \tau''$. Since the similarity functions are clearly equicontinuous in τ'', we conclude by Theorem 2.10 that $\tau'' = \tau'$.

3.2. The vector space model of Everett and Cater (1992)

We briefly recall the structure of the vector space model as presented in (Everett & Cater, 1992) and point out some problems with this model. We will compare this vector model to another one in Section 3.3.

The Everett-Cater model starts by taking $I^n = [0,1]^n$ for the document space as well as far the query space with the following similarity function :

$$\text{sim}(D,Q) = \frac{< D,Q >}{\|D\|_2 \|Q\|_2} = \cos(D,Q) \tag{13}$$

where $<.,.>$ denotes the usual inner product, $\|.\|_2$ is the Euclidean norm and cos(D,Q) denotes the cosine of the angle between the lines OD and OQ.

We see two problems with this model. Firstly, the model does not distinguish between documents on a straight line through the origin. Since documents in $[0,1]^n$ are often obtained through a weighing process, it seems to us a waste of possibilities not to use these different weights. Secondly, the similarity functions are not defined in points (D,Q) where at least one of the coordinates is zero. Moreover, it is impossible to extend sim in such a way that this function becomes continuous on I^n.

Everett and Cater solve these problems by taking DS = QS = $I^n \setminus \{0\}$ and introducing an equivalence relation R, where DRE iff there is straight line through the origin containing both D and E. Then DS* = $(I^n \setminus \{0\})/R$ with the usual quotient norm (and hence quotient topology). For readers not familiar with quotient theory we just note here that elements of DS* are straight half lines through the origin (0 not included) and only the part inside I^n is used. The similarity function (13) can be used unambiguously as above : if $D^* \in DS^*$ and $Q \in QS$ then

$$\text{sim } (D^*, Q) = \frac{< D, Q >}{\|D\|_2 \|Q\|_2} \tag{14}$$

for any representative $D \in D^*$. Also, for $D^*, E^* \in DS^*$, we can define :

$$\text{sim } (D^*, E^*) = \frac{< D, E >}{\|D\|_2 \|E\|_2} \tag{15}$$

for any representative $D \in D^*$, $E \in E^*$.

By taking quotients as above we now obtain that the retrieval system separates points. Indeed, if sim(D*,Q) = sim(E*,Q), $\forall Q \in$ QS, then cos(D,Q) = cos(E,Q) for any $D \in D^*$, $E \in E^*$, and hence $D^* = E^*$. This shows that here τ' is a metric topology. This leads to:

Theorem 3.1 (Everett & Cater, 1992) :

$(DS^*, \tau') = (S, \mathcal{E})$ with S that part of the unit sphere that contains vectors with positive (or zero) coordinates and with $\mathcal{E}$ the Euclidean topology on S.

Proof :

The homeomorphism is obtained via the function :

$$f : DS^* \rightarrow S : D^* \rightarrow D \cap S \qquad \square \tag{16}$$

Theorem 3.2 :

$\tau = \tau'' = \tau'$ on DS^* and on DS.

Proof :

The subbasic neighborhoods for an element $D \in DS$ or $D^* \in DS^*$, for τ as well as for τ', are of the form : an open cone with top 0 (not an element of DS^*) and D^* as central ray. Hence $\tau = \tau'$, and thus, by Theorem 2.5, $\tau = \tau' = \tau''$. □

Corollary 3.3 :

DS^* with $\tau = \tau' = \tau''$ is compact.

Proof :

This follows from Theorem 3.1 and Theorem 3.2.

Corollary 3.4 (Everett & Cater, 1992) :

DS with $\tau = \tau' = \tau''$ is compact.

Proof :

The argument that proves that $(S,\mathscr{E})$ is compact can be used on DS if we use the function f^{-1} (the inverse of function f of (16)) and consider D^* as a subset of DS. □

Note.

It is not true that DS and DS^* are homeomorphic! Indeed, DS^* is a Hausdorff (or T_2) space, while DS is not.

The vector space model that we will present in the next section is an alternative for the usual one. Its mathematical properties are less complicated and more interesting as it will be defined on I^n (not on a quotient space) and will give different topologies $\tau \neq \tau''$. In addition to this it takes into account the different weights of vectors in I^n .

3.3. An alternative for the classical vector space model

Using the cosine of the angle between vectors does not take into account the different weights given to each coordinate (representing e.g. a keyword). In addition to this, documents can be situated close to each other, yet the similarity as measured by cosines can be relatively small too. Using the inner product between vectors takes care of these problems. In such a model a query gives weights to keywords. These weights can be intepreted as minimum requirements : documents with higher weights (for that particular keyword) will just score better.

We will now formalize this. Let $DS = QS = I^n = [0,1]^n$, $n \in \mathbf{N}_0$. For every $D \in DS$, $Q \in QS$:

$$\text{sim}(D,Q) = < D,Q > = \|D\|_2 \|Q\|_2 \cos(D,Q) \qquad (17)$$

The following results describe the topologies of this retrieval model.

Theorem 3.5 :

(i) The function $<.,.> : DS \times QS \rightarrow \mathbf{R}^+; (D,Q) \rightarrow <D,Q>$ is continuous for the norm topologies on $DS = QS = I^n$.

(ii) The retrieval system (DS, QS = DS, sim), with sim defined in (17) separates the points of DS.

The proof of these results is well-known and is omitted.

Therem 3.6 :

In this model $\tau' = \mathscr{E}$, the Euclidean topology on I^n. Hence, here DS with any of the topologies τ, τ' or τ'' is compact.

Theorem 3.7 :

For this model, $\tau'' = \tau'$, hence DS with this topology is a T_2 space.

Theorem 3.8 :

For this model, (DS, τ) is a T_0-space that is not a T_1-space. Consequently $\tau \neq \tau''$.

The proof of these theorems can be found in Appendix D.

3.4. Another simple example of a document system for which $\tau \neq \tau'' = \tau'$

Let DS = {a,b,c,d}, QS = {e} and define

sim(a,e) = 0.5 = sim(d,e)

sim(b,e) = 1, sim(c,e) = 0

Then :

τ = {ϕ, D S, {a, b, d }, {b }}

while

$\tau' = \tau''$ = {ϕ, DS, {b}, {c}, {a,d}, {b,c}, {a, b, d}, {a,c,d}} $\neq$ (DS) ,

the discrete topology. Note that the equality between τ' and τ'' illustrates Corollary 2.11, as here QS is a finite set. This retrieval system does not separate points.

3.5. Three examples for which $\tau = \tau' = \tau'' = \rho$(DS)

Example 3.2 gave an example for which $\tau = \tau' = \tau''$, but different from the discrete topology since they were homeomorphic with the Ecuclidean topology on S. Three examples will follow where all topologies are equal to ρ(DS), the discrete topology on DS.

3.5.1.

DS = QS any set

$$\text{sim}(D,Q) \begin{cases} = 0, & D \neq Q \\ = 1, & D = Q \end{cases} \qquad (18)$$

Hence in this IR-model only exact matches are allowed!

So $\forall Q \in QS, \forall r \in \mathbf{R}, R(Q,r) = \phi$ if $r \geq 1$ and $= \{Q\}$ if $r < 1$.

Hence $\tau = \wp(DS)$. By theorem 2.5, $\tau \subset \tau'' \subset \tau'$. Hence $\tau = \tau'' = \tau' = \wp(DS)$.

Note that the metric of τ' is the discrete metric

$$d(D,E) \begin{cases} = 1 & \text{if } D = E \\ = 0 & \text{if } D \neq E \end{cases} \tag{19}$$

Discrete metrics yield discrete topologies but not conversely as the next two examples show.

3.5.2.

The next example deals with documents that are ordered according to "similarity". This is an index approach and can not always be used in practice. However, if topics can be ordered this way (e.g. for aspects on which a linear order applies such as distance, temperature, ...) the example is useful. Let

$$DS = \{D_1,\ldots,D_n\} = QS, \quad n \in \mathbf{N} \quad \text{fixed}$$

Define

$$\text{sim}\ (D_i, D_j) = \frac{i + j}{2\ \max\ (i, j)} \tag{20}$$

Intuitively this means that if i and j are far apart then $\text{sim}(D_i,D_j)$ is small and if $i \approx j$, then $\text{sim}(D_i,D_j) \approx 1$. In any case is $\text{sim}(D_i,D_j) = 1$, for every $i \in \{1,\ldots,n\}$. We have the following results :

Theorem 3.9 :

$\forall j = 1,\ldots,n, \forall r > 0.$

$$R(D_j,r) = \left\{ D_i \in DS \,\middle|\, i \in \{1,\ldots,n\} \cap \left]2jr - j,\ \frac{j}{2r - 1}\right[\right\} \tag{21}$$

which forms a subbasis for τ.

Theorem 3.10 :

Let d be the metric of τ'. Then,$\forall i,k = 1,\ldots,n$:

$$d(D_i, D_k) = \frac{|i - k|}{2 \max(i,k)} \tag{22}$$

Corollary 3.1.1 :

$\tau = \tau'' = \tau' = \rho(DS)$.

For the proofs we refer the reader to Appendix E.

A similar example, but with a completely different similarity function, now follows.

3.5.3.

$$DS = QS = \{D_1,\ldots,D_n\}, \quad n \in \mathbf{N} \quad \text{fixed}$$

Define

$$\text{sim}(D_i, D_j) = \frac{2}{\pi} \text{Arctan}\left(\frac{1}{|i - j|}\right) \tag{23}$$

Note again that $\text{sim}(D_i, D_i) = 1$, $\forall i = 1,\ldots,n$

Theorem 3.12 :

$\forall j = 1,\ldots,n, \forall r > 0.$

$$R(D_j, r) = \left\{ D_i \in DS \middle| i \in \{1,\ldots,n\} \cap \left] j - \frac{1}{\tan\frac{\pi r}{2}}, j + \frac{1}{\tan\frac{\pi r}{2}} \right[\right\} \tag{24}$$

which forms a subbasis for τ.

Theorem 3.13 :

Let d be the metric of τ'. Then,$\forall i,k = 1,\ldots,n$:

$$d(D_i, D_k) = \frac{2}{\pi} \text{Arctan}\,|i - k| \tag{25}$$

Corollary 3.14 :

$\tau = \tau'' = \tau' = \rho(DS)$

For the proofs we refer the reader to Appendix F.

3.6. An example for which $\tau \neq \tau' \neq \tau''$

This final example shows that $\tau \neq \tau' \neq \tau''$ is possible. In view of Theorem 2.10 and its Corollary 2.11, QS must be an infinite space. This means that we must admit an infinite number of possible queries.

Let DS = QS = $[0,+\infty[$, with sim(D,Q) = D.Q (the simple product of the numbers D and Q). Let $(D_n)_{n=1,\ldots,\infty}$ be a strictly increasing sequence in DS, converging to D in the usual, Euclidean norm on $[0,+\infty[$ (i.e. the absolute value). Consequently, $\lim_{n\to\infty} D_n = D$ in τ'' since, for every $Q \in QS$:

$$\text{sim } (D_n, Q) = D_n.Q \to D.Q = \text{sim } (D, Q)$$

Now, D_n does not tend to D in τ' since

$$d'(D_n, D) = \sup_{Q \in QS} (\min (1, |\text{sim } (D_n, Q) - \text{sim } (D, Q)|))$$

which is equal to 1 if $D_n \neq D$ and is equal to 0 if $D_n = D$. As here D_n is never equal to D, $d'(D_n D)$ is always equal to 1.

From this it follows that $\tau' = \rho(DS) \neq \tau''$. We still have to prove that $\tau \neq \tau''$. We know already that τ'' yiels a T_2-space, since τ'' coincides here with the Euclidean topology on $[0,+\infty[$. (This follows from the fact that a sequene $D_n \to D$ in $|.|$ iff $D_n.Q \to D.Q$ for every $Q \in [0,+\infty[$.) Now, (DS, τ) is not even a T_1-space : for every $D \in DS$ and $E = \alpha D$, $\alpha > 1$, and for every τ-neighborhood U of D, $\exists n \in \mathbf{N}$ and $R(F_j, r_j)$ such that

$$D \in \bigcap_{j=1}^{n} R(F_j, r_j) \subset U$$

Hence $sim(D,F_j) = D.F_j > r_j$, for every $j = 1,\ldots,n$. Since $D.F_j \in [0,+\infty[$ also $sim(E,F_j) = \alpha D.F_j > r_j$, which implies that $E \in U$. This shows that (DS,τ) is not a T_1-space.

In fact, if $[0,+\infty[$ is equiped with the topology D inherited from S on $\mathbf{R}$ (cf. Theorem 2.9), then $\tau = D$ (restricted to $[0,+\infty[$). We conclude :

τ is the topology inherited from S on $\mathbf{R}$

τ" is the Euclidean topology

τ' is the discrete topology on DS,

and all three of them are different.

4. Boolean information retrieval as a subsystem of any topological IR-system

In Cater (1986) and Everett and Cater (1992) the Boolean IR-model is shown to be an example of topological retrieval, in the sense that specially defined similarity values yield traditional Boolean retrieval. In this section we will show that Boolean retrieval can be introduced in any IR-model (DS, QS, sim) without having to specify the value of a similarity function for Boolean combinations of elementary queries. The Boolean model we will present has the further advantage that it presents a major clarification why the introduction of topologies in theoretical IR-models is essential.

As defined in Section 1, let, for every $Q \in QS$,

$$ret_\tau(Q) = \{R(Q,r) \mid r \in \mathbf{R}\}$$

and

$$ret_{\tau''}(Q) = \{U(Q,r_1,r_2) \mid r_1,r_2 \in \mathbf{R},\ r_1 < r_2\}$$

be the retrieval sets of Q w.r.t. the topology τ (resp. τ"). In the sequel of this section we will focus on $ret_\tau(Q)$, denoting this set simply as ret(Q). (Exactly the same reasoning can be made for τ").

The set QS consists of elementary queries, which means that usually logical combinations of queries do not belong to QS although they are not excluded either. ret(Q) consists of all possible results when Q is used as a query. The only difference between the use of the topologies τ and τ'' is that we require 'minimum' (threshold) values for sim (τ case), or 'close' values (τ'' case).

Recall also that the sets in all $ret_\tau(Q)$, $Q \in QS$, form a subbasis for τ. This means that

$$B = \left\{ \bigcap_{i=1}^{n} R(Q_i, r_i) \mid Q_i \in QS, r_i \in \mathbb{R},\ n \in \mathbb{N}_0 \right\}$$

is a basis for τ (similarly a basis for τ'' can be built using the sets $U(Q, r_1, r_2)$). An open set A in τ is then any union (finite or infinite) of sets in B :

$$A = \bigcup_{j \in J} \left(\bigcap_{i=1}^{n_j} R(Q_{ij}, r_{ij}) \right)$$

where J is a finite or infinite index set. From now on we will assume that DS and QS are finite. Note that then $\tau'' = \tau'$ (Theorem 2.10 and Corollary 2.11). So we have only τ and τ'' to consider, and, as stated before, we will focus on τ. In this case a general open set A in τ has the form

$$A = \bigcup_{j=1}^{m} \left(\bigcap_{i=1}^{n_j} R(Q_{ij}, r_{ij}) \right) \tag{26}$$

Definition 4.1 :

Let $Q_1, \ldots, Q_n$ be n elements of QS. We introduce, symbolically, an element $\bigotimes_{i=1}^{n} Q_i$ and define

$$ret\left(\bigotimes_{i=1}^{n} Q_i \right) = \left(\bigcap_{i=1}^{n} R(Q_i, r_i) \mid r_i \in \mathbf{R} \right) \tag{27}$$

We call $\overset{n}{\underset{i=1}{\otimes}} Q_i$ the Boolean AND applied to the elementary queries $Q_1,\ldots,Q_n \in QS$. By definition, this Boolean AND retrieves sets in $ret\left(\overset{n}{\underset{i=1}{\otimes}} Q_i\right)$. We repeat that $\overset{n}{\underset{i=1}{\otimes}} Q_i$ is not necessarily an element of QS. Similarly we define a Boolean OR.

Definition 4.2 :

Let $Q_1,\ldots,Q_n$ be n elements of QS. Consider, symbolically, an element $\overset{n}{\underset{i=1}{\oplus}} Q_i$ and define

$$ret\left(\overset{m}{\underset{j=1}{\oplus}} Q_j\right) = \left\{\overset{m}{\underset{j=1}{\cup}} R(Q_j,r_j) \mid r_j \in \mathbb{R}\right\} \tag{28}$$

The set $\overset{m}{\underset{j=1}{\oplus}} Q_j$ is called the Boolean OR applied to the elementary queries $Q_1,\ldots,Q_n \in$ QS. By definition, this Boolean OR retrieves sets in $ret\left(\overset{m}{\underset{j=1}{\oplus}} Q_j\right)$.

Of course, we can as well define Boolean ANDs and Boolean ORs with respect to the topology τ". We suppose that it is clear in every IR-search whether we want to work with τ (thresholds) or τ" (close matches).

Based on Definitions 4.1 and 4.2 we can easily define more arbitrary Boolean queries, based on elementary queries in QS.

Definition 4.3 :

Let $(Q_{ij})_{i=1,\,j=1}^{n\quad m}$ be an array of queries in QS.

The Boolean query (not necessarily in QS) $\overset{m}{\underset{j=1}{\oplus}}\left(\overset{n_j}{\underset{i=1}{\otimes}} Q_{ij}\right)$ is defined through its retrieval:

$$\text{ret}\left(\bigoplus_{j=1}^{m}\left(\bigotimes_{i=1}^{n_j} Q_{ij}\right)\right) = \left\{\bigcup_{j=1}^{m}\left(\bigcap_{i=1}^{n_j} R(Q_{ij}, r_{ij})\right) \mid r_{ij} \in \mathbb{R}\right\} \tag{29}$$

and similarly when $\oplus$ and $\otimes$ (hence $\cup$ and $\cap$) are interchanged.

We now need the following lemma :

Lemma 4.4 : (Dugundji (1966), p.25)

Let $\{B_\alpha \mid \alpha \in A\}$ be a family of sets and assume that $\{A_\lambda ; \lambda \in \Lambda\}$ is a partition of A (hence each $A_\lambda \neq \phi$). Let $T = \coprod_{\lambda\in\Lambda} A_\lambda$.

Then

$$\bigcap_{\lambda\in\Lambda}\left(\bigcup_{\alpha\in A_\lambda} B_\alpha\right) = \bigcup_{t\in T}\left(\bigcap_{\lambda\in\Lambda} B_{t(\lambda)}\right) \tag{30}$$

where $t(\lambda) \in A_\lambda$ and $t = (t(\lambda))_{\lambda=\Lambda}$.

By taking complements one also has :

$$\bigcup_{\lambda\in\Lambda}\left(\bigcap_{\alpha\in A_\lambda} B_\alpha\right) = \bigcap_{t\in T}\left(\bigcup_{\lambda\in\Lambda} B_{t(\lambda)}\right) \tag{31}$$

In plain terms : any introduction of unions of sets A_{ij} can be interpreted as a union of intersections of the same sets (but in another order) and vice-versa.

We have now reached the following important result, yielding a major reason why topologies on IR-systems are useful.

Theorem 4.5 :

For any IR-model (DS, QS, sim) with finite DS and QS we have the following equalities:

a) the topology τ is equal to the set of all possible Boolean retrievals based on elementary queries Q in QS, using thresholds;

b) the topology τ" is equal to the set of all possible Boolean retrievals based on elementary queries Q in QS, using close matches.

Proof :

The principal elements of the proof have already been outlined. An arbitrary Boolean query can be denoted as a combination of AND and OR Boolean queries in any order, which can be denoted as in (29) or with $\cup$ and $\cap$ interchanged (cf. Definition 4.3). By lemma 4.4 its set of retrievals can be written in the form

$$\left\{ \bigcup_{j=1}^{m} \left(\bigcap_{i=1}^{n_j} \mathbf{R}\left(Q_{ij}, r_{ij}\right) \right) \middle| r_{ij} \in \mathbf{R} \right\}$$

Letting m and n_j vary over all possible natural numbers and $Q_{ij} \in$ QS, we see, by (26), that the set of all Boolean retrievals using thresholds, is nothing but the topology τ. Similarly, the set of all Boolean retrievals using close matches is nothing but the topology τ".

The following result gives a relation between the usual OR relation and our somewhat more general use of the Boolean OR.

Proposition 4.6 :

If Q_1 and Q_2 are single attribute queries, if also $Q = Q_1$ OR Q_2 belongs to QS, and if sim(D,Q) is defined as max(sim(D,Q_1), sim(D,Q_2)) - as in the classical Boolean or in the fuzzy set case - then (using thresholds)

$$\text{ret}(Q) \subset \text{ret}\left(\bigoplus_{i=1}^{2} Q_i \right)$$

Proof :

Consider the set {D $\in$ DS | sim(D,Q) > r}, an element of ret(Q). As sim(D,Q) = max(D,Q_1), sim(D,Q_2)), this set is equal to {D $\in$ DS | max(sim(D,Q_1), sim(D,Q_2)) > r}

= {D $\in$ DS | sim(D,Q_1) > r, or, sim(D,Q_2)) > r}

= {D $\in$ DS | sim(D,Q_1) > r} $\cup$ {D $\in$ DS | sim(D,Q_2) > r}

= R(Q_1,r) $\cup$ R(Q_2,r) $\in$ ret($Q_1 \oplus Q_2$).

A similar result can be shown for any finite disjunctive form and for any finite conjunctive form. The Boolean NOT will be studied in a following article.

5. SUMMARY

In this article Everett and Cater's retrieval and pseudo-metric topologies are examined. Counterexamples to two statements are given and results correcting the original statements are presented. These corrections lead to the introduction of a new topology, namely the similarity topology. This new topology satisfies the interesting property of making the similarity functions continuous (stable).

Several examples are presented amongst them a modified vector space model. Finally, the article shows that the retrieval and the similarity topology can be considered as the sets of retrievals of arbitrary Boolean AND-OR queries.

REFERENCES

CATER, S.C. (1986). The topological information retrieval system and the topological paradigm : a unification of the major models of information retrieval. Dissertation, Louisiana State University, Baton Rouge, LA.

CSASZAR, A. (1978). General topology. Disquisitiones Mathematicae Hungaricae 9. Budapest : Akadémiai Kiadó.

DUGUNDJI, J. (1966). Topology. Boston : Allyn and Bacon.

EGGHE, L. and ROUSSEAU, R. (1997 a). Duality in information retrieval and the hypergeometric distribution. Journal of Documentation, 53, 488-496, 1997.

EGGHE, L. and ROUSSEAU, R. (1997 b). Everett and Cater's retrieval topology (letter to the editor). Journal of the Amerian Society for Information Science, 48, 479-481, 1997.

EVERETT, D.M. and CATER, S.C. (1992). Topology of document retrieval systems. Journal of the American Society for Information Science, 43, 658-673.

KREYSZIG, E. (1978). Introductory functional analysis with applications. New York : Wiley & Sons.

SALTON, G. and McGILL, M.J. (1983). Introduction to modern information retrieval. New York : McGraw-Hill.

STEEN, L.A. and SEEBACH Jr, J.A. (1978). Counterexamples in topology. Berlin : Springer Verlag.

WILANSKY, A. (1970). Topology for analysis. Lexington (MA) : Xerox College.

WILLARD, S. (1970). General topology. Reading (MA) : Addison-Wesley.

APPENDIX A
GENERALITIES ON TOPOLOGICAL SPACES

Let X denote a set. Denote by $\mathcal{P}(X)$ the set of all subsets of X. A topology τ on X is a subset of $\mathcal{P}(X)$ such that

(i) any union of elements in τ belongs to τ

(ii) any finite intersection of elements in τ belongs to τ

(iii) ϕ (the empty set) and X belong to τ.

The elements of τ are called open sets.

The couple (X,τ) is called a topological space.

Given two topologies τ_1 and τ_2 on X we say that τ_1 is weaker (or smaller or coarser) than τ_2 if $\tau_1 \subset \tau_2$. We then also say that τ_2 is stronger (larger or finer) than τ_1.

A set $F \subset X$ is called closed if its complement $F^c \in \tau$. If $A \subset X$, the closure of A in X, w.r.t. τ, denoted by $\bar{A}$, is the set

$$\overline{\mathbf{A}} = \bigcap \{\mathbf{B} \subset \mathbf{X} | \mathbf{B} \text{ is closed and } \mathbf{A} \subset \mathbf{B}\}$$

If it is not clear that the closure is w.r.t. τ we denote $\bar{A}^{\tau}$.

If (X,τ) is a topological space, a base for τ is a collection $B \subset \tau$ such that

$$\tau = \{ \bigcup_{B \in C} \mathbf{B} | C \subset B \}$$

A subbase for τ is a collection $C \subset \tau$ such that the collection of all finite intersections of elements of C forms a base for τ. Any collection of subsets of a set X is a subbase for some topology on X. This follows from the definition of τ. This assertion is not true for a base!

Note that B is a basis for τ iff, whenever $G \in \tau$ and $x \in G$, there is a $B \in B$ such that $x \in B \subset G$.

Let $x \in X$. We say that $U \subset X$ is a neighborhood of x if there exists $G \in \tau$ such that $x \in G \subset U$. We denote by $V_\tau(x)$ the collection of all neighborhoods of $x \in X$ and it is called a neighborhood system. A neighborhood base at $x \in X$ is a collection $B_\tau(x) \subset V_\tau(x)$ such that each $U \in V_\tau(x)$ contains an element $V \in B_\tau(x)$, i.e. $V_\tau(x)$ is determined by B_τ as

$$V_\tau(x) = \{U \subset X | V \subset U, \exists V \in B_\tau(X)\}$$

The elements of $B_\tau(x)$, once chosen, are called basic neighborhoods.

Let (X,τ) be a topological space. We say that it is a T_0 space if for every $x,y \in X$, $x \neq y$, there exists a neighborhood of one of these points not containing the other. We say that it is a T_1 space if for every $x,y \in X$, $x \neq y$, there exist neighborhoods of each of these points not containing the other point.

Let (X,τ) be a topological space. We say that (X,τ) is a Hausdorff (or T_2) space if for every $x,y \in X$, $x \neq y$, there exists $G,H \in \tau$ such that $x \in G$, $y \in H$ and $G \cap H = \phi$. Equivalently if there exists $U \in V_\tau(x)$, $V \in V_\tau(y)$ such that $U \cap V = \phi$.

Clearly $T_2 \Rightarrow T_1 \Rightarrow T_0$. In this paper we will have examples showing that the converse is not true.

Let X be a set and

$$d : X \times X \to \mathbf{R}^+$$
$$(x,y) \mapsto d(x,y)$$

a function such that

(i) $d(x,x) = 0$

(ii) $d(x,y) = d(y,x)$

(iii) $d(x,y) \leq d(x,z) + d(z,y)$ (triangle inequality)

for every $x,y,z \in X$. Then d is called a pseudo-metric on X and (X,d) is called a pseudo-metric space. Every pseudo-metric space can be considered as a topological space as follows :

(a) Define the "open balls", $\forall x \in X, \forall \varepsilon > 0$

$$\mathbf{B(x,\epsilon) = \{y \in X | d(x,y) < \epsilon\}}$$

(b) Define sets $A \subset X$ open iff for every $x \in A$ there is a $B(x,\varepsilon) \subset A$.

It can be shown (cf. Willard (1970), 2.6, p.18) that the open sets defined above form a topology on X, the so-called pseudo-metric topology on X.

If (i) above can be reformulated as

(i)' $d(x,y) = 0$ iff $x = y$

then we call d a metric, (X,d) a metric space and the topology that it generates, the metric topology.

If X is a vector space over $\mathbf{R}$ we can even go further. We say that

$$\|.\| : \mathbf{X} \to \mathbf{R}^+$$

is a pseudo-norm on X if

(i) $\|\vec{0}\| = 0$ ($\vec{0}$ is the zero vector in X)

(ii) $\|\alpha x\| = |\alpha| \, \|x\|$, $\forall \alpha \in \mathbf{R}$

(iii) $\|x + y\| \leq \|x\| + \|y\|$

for all $x,y \in X$. $\|.\|$ is called a norm if (i) can be replaced by

(i)' $\|x\| = 0$ iff $x = \vec{0}$

for all $x \in X$.

X, $\|.\|$, accordingly, is called a (pseudo-)normed space. If we define

$$\mathbf{d(x,y) = \|x - y\|}$$

then it can be shown that d is a (pseudo-)metric.

Example :

$\mathbf{R}^n$, $\|.\|$, with, if $x = (x_1,\ldots,x_n) \in \mathbf{R}^n$

$$\|x\|_p = \left(\sum_{i=1}^{n} |x_i|^p \right)^{1/p} ,$$

for $p \geq 1$.

$\|.\|_p$ is called the Minkowski norm and, if $p = 2$, $\|.\|_2$ is called the Euclidean norm. d derived from $\|.\|_2$ is called the Euclidean metric and its topology the Euclidean topology, denoted by $\mathcal{E}$. For $p = \infty$ we define

$$\|x\|_\infty = \max_{i=1,\ldots,n} |x_i|$$

Also $\|.\|_\infty$ is a norm on $\mathbf{R}^n$. All norms $\|.\|_p$ $(1 \leq p \leq \infty)$ are equivalent by which we mean that they all induce the same topology (in this case $\mathcal{E}$). The same definition goes for equivalent pseudo-metrics.

$\|.\|_2$ has one special feature, however. Define, for $x,y \in \mathbf{R}^n$

$$\langle x,y \rangle = \sum_{i=1}^{n} x_i y_i$$

the so-called inproduct in $\mathbf{R}^n$. This inproduct is used in this article. It makes $\mathbf{R}^n$ into an inproduct space. Every inproduct space X is a normed space via the definition

$$\|x\| = \sqrt{\langle x,x \rangle}$$

for every $x \in X$.

The inproduct satisfies the so-called inequality of Cauchy-Schwarz : $\forall x,y \in X$

$$|\langle x,y \rangle| \leq \|x\|.\|y\|$$

Let (X,τ) be a topological space and $Y \subset X$.
Then

$$\tau_Y = \{G \cap Y | G \in \tau\}$$

is a topology on Y, the subspace topology induced on Y by τ.

Example :

$I^n \subset \mathbf{R}^n$ where $I = [0,1]$.

We can take the restriction of $\|.\|_2$ (or any other norm on $\mathbf{R}^n$) to I^n. This yields the Euclidean topology on I^n, being the subspace topology of I^n, inherited from the Euclidean topology on $\mathbf{R}^n$.

Let (X,τ) and (Y,τ') be two topological spaces and let $f : X \to Y$ be a function. We say that f is continuous if for every $B \in \tau'$, $f^{-1}(B) \in \tau$. We say that f is open if $G \in \tau$ implies $f(G) \in \tau'$. If $f : X \to Y$ is a bijection that is open and continuous then we say that (X,τ) and (Y,τ') are homeomorphic and we denote $(X,\tau) \cong (Y,\tau')$. Topologically spoken these spaces are indistinguishable.

Let (X,τ) be a topological space. We say that τ is compact if every open cover of X has a finite subcover.

Example :

$[0,1]$ is compact but $]0,1[$ is not ($]1/n,1[$, $n \in \mathbf{N}$ is an open cover of $]0,1[$ without a finite subcover).

A set in $\mathbf{R}^n$ (with the Euclidean topology $\mathcal{E}$) is compact iff it is closed and bounded.

Let $f_j : (X,\tau) \to \mathbf{R}$ be a family of functions ($j \in J$, where the index set J can be any, denumerable or non-denumerable, set). We say that the family $(f_j)_{j\in J}$ forms an equicontinuous family in $x \in X$, if for every $\varepsilon > 0$, there is a neighborhood U of x, independent of $j \in J$, such that $y \in U$ implies that for every $j \in J$: $|f_j(x) - f_j(y)| < \varepsilon$. The family $(f_j)_{j\in J}$ is said to equicontinuous on (X,τ) iff it is equicontinuous in every point x of X. Note that every finite family of continuous functions is an equicontinuous family. Moreover, if the topology $\tau = \mathcal{P}(X)$, the discrete topology, then any family is equicontinuous.

APPENDIX B
PROOF OF THEOREM 2.1

Theorem 2.1 :

Let d, d′ and d" be as defined in (9), (11) and (12). Then

(i) the (pseudo-)metrics d′ and d" are equivalent

(ii) if d exists, it is equivalent with d′ and d".

Proof :

For every $\varepsilon > 0$ and $D \in DS$, we denote by $B(D,\varepsilon)$, $B'(D,\varepsilon)$ and $B"(D,\varepsilon)$ the d- (resp. d′ and d") ε-open ball around D.

(i) It is clear that for every $D,E \in DS$, $d'(D,E) \leq d"(D,E)$, hence $B"(D,\varepsilon) \subset B'(D,\varepsilon)$ for every $\varepsilon > 0$ and $D \in DS$. Let now $\varepsilon > 0$ and $D \in DS$ be given. Set $\varepsilon' = \varepsilon/(2+\varepsilon)$. Then $d'(D,E) < \varepsilon'$ implies

$$\frac{\rho_Q(D,E)}{1 + \rho_Q(D,E)} < \varepsilon'$$

for every $Q \in QS$. Hence, $\rho_Q(D,E) < \varepsilon/2$ for every $Q \in QS$. So, $d"(D,E) < \varepsilon$ and hence $B'(D,\varepsilon') \subset B"(D,\varepsilon)$. This shows that d′ and d" generate the same topology on DS, denoted as τ'.

(ii) Assume now that d exists. Since for every $D,E \in DS$: $d'(D,E) \leq d(D,E)$, we see that $B(D,\varepsilon) \subset B'(D,\varepsilon)$ for every $\varepsilon > 0$ and $D \in DS$. Let now $B(D,\varepsilon)$ be an arbitrary open ε-ball around D. Let $\varepsilon' = \varepsilon/(2+\varepsilon) > 0$, then we have for every $E \in B'(D,\varepsilon')$ that

$$d'(D,E) = \sup_{Q \in QS} \frac{\rho_Q(D,E)}{1 + \rho_Q(D,E)} < \varepsilon'$$

This implies that, for every $Q \in QS$, $\rho_Q(D,E) < \varepsilon/2$. Hence, $d(D,E) < \varepsilon$, which shows that $B'(D,\varepsilon) \subset B(D,\varepsilon)$. This proves part (ii) of the theorem. □

APPENDIX C
PROOF OF THEOREM 2.10

Theorem 2.10 :

The following assertions are equivalent :

(i) $\tau' = \tau''$

(ii) The family $\{sim(.,Q) | Q \in QS\}$ is equicontinuous on (DS,τ''), where the functions range in $R, \mathcal{E}$.

Proof : (ii) $\Rightarrow$ (i) :

Let $(D_i)_{i\in I}$ be a net in DS, convergent to $D \in DS$ for τ''. Since the set $\{sim(.,Q) | Q \in QS\}$ is equicontinuous on τ'' into $\mathbf{R}$, $\mathcal{E}$, we have that $\forall \varepsilon > 0$, $\exists i_0(\varepsilon) \in I$ (independent of $Q \in QS$) such that $\forall i \geq i_0(\varepsilon)$

$$|sim(D_i,Q) - sim(D,Q)| < \frac{\epsilon}{2}$$

$\forall Q \in QS$. Hence, $\forall Q \in QS$

$$\frac{|sim(D_i,Q) - sim(D,Q)|}{1 + |sim(D_i,Q) - sim(D,Q)|} < \frac{\epsilon}{2}$$

Hence

$$\sup_{Q\in QS} \frac{|sim(D_i,Q) - sim(D,Q)|}{1 + |sim(D_i,Q) - sim(D,Q)|} < \epsilon$$

so $d'(D_i,D) < \varepsilon$. Consequently : $(D_i)_{i\in I}$ converges to D in τ' (by using the definition of τ'). Hence $\tau' = \tau''$.

(i) $\Rightarrow$ (ii) :

Let the net $(D_i)_{i\in I}$ be convergent to D on DS equiped with the topology $\tau'' = \tau'$. Hence $\forall \varepsilon > 0$, $\exists i_0(\varepsilon) \in I$ such that $\forall i \geq i_0(\varepsilon)$, $i \in I$:

$$d'(D_i,D) < \frac{\epsilon}{1+\epsilon}$$

From this it follows that, $\forall Q \in QS$

$$|sim(D_i,Q) - sim(D,Q)| < \epsilon$$

(hence $\forall i \geq i_0(\epsilon)$, independent of Q). Since this goes for every net $(D_i)_{i \in I}$ that converges in τ'' we have that the set

$$\{sim(.,Q) | Q \in QS\}$$

is equicontinuous on (DS,τ''). □

APPENDIX D
PROOF OF THEOREMS 3.6, 3.7 AND 3.8

Theorem 3.6 :

$\tau' = \mathscr{E}$. Hence τ, τ'', τ' are compact.

Proof :

Let $|||.|||$ be defined by, for $D \in DS$

$$|||D||| = \sup_{|Q|_\infty \leq 1} |<Q,D>|$$

$$= \sup_{Q \in QS} |<Q,D>|$$

(since $D \in I^n$ we can indeed restrict ourselves to vectors Q with positive coordinates). Hence d defined by

$$d(D,E) = |||D - E|||$$

yields the metric topology τ'. Furthermore d' defined by

$$d'(D,E) = \|D - E\|_2$$

with

$$\|D\|_2 = \sup_{|Q|_2 \leq 1} |<Q,D>|$$

yields the Euclidean metric. Since $\|.\|_\infty$ and $\|.\|_2$ are equivalent (see Appendix A) and since $<\alpha Q,D> = \alpha <Q,D>$ for every $\alpha \in \mathbb{R}$, it follows that d and d' are equivalent and hence $\tau' = \mathscr{E}$.

Since $\tau \subset \tau'' \subset \tau' = \mathscr{E}$ and since $\mathscr{E}$ is compact, also τ, τ'' and τ' are. □

Theorem 3.7 :

$\tau' = \tau''$, hence they are T_2-spaces.

Proof :

In view of theorem 2.10 it suffices to prove that the set

$$\{<.,Q> | Q \in QS\}$$

is equicontinuous on (I^n,τ''). For this it suffices to prove that

$$\sup_{Q \in QS} \|<.,Q>\| < \infty$$

(see e.g. Wilansky (1970), p.291, ex.201).

Now

$$\sup_{Q \in QS} \|<.,Q>\|$$

$$= \sup_{Q \in QS} \sup_{\|D\|_2 \leq 1} |<D,Q>| \leq \sup_{Q \in QS} \sup_{\|D\|_2 \leq 1} \|D\|_2 \, \|Q\|_2 \quad \text{(Cauchy-Schwarz)}$$

$$= \sup_{\|D\|_\infty \leq 1} \sup_{\|D\|_2 \leq 1} \|D\|_2 \, \|Q\|_2 \quad .$$

Since $\|.\|_\infty$ and $\|.\|_2$ are equivalent, the result follows. From theorem 3.6 it now follows that $\tau' = \tau''$ are T_2-spaces (since $\mathcal{E}$ is). □

Theorem 3.8 :

τ is a T_0-space but not a T_1-space. Hence $\tau \neq \tau''$.

Proof : (i) τ is T_0

Let $D,E \in DS$, $D \neq E$ and we are looking for (sufficient condition)

$$[D \in R(D,r) \wedge E \notin R(D,r)]$$

$\vee$

$$[D \notin R(E,r) \wedge E \in R(E,r)]$$

Hence we look for $D = (D_1,...,D_n)$, $E = (E_1,...,E_n)$ in DS such that

$$\left[\langle D,D \rangle = \sum_{i=1}^{n} D_i^2 > r \wedge \langle D,E \rangle = \sum_{i=1}^{n} D_i E_i \leq r \right]$$

v

$$\left[\langle E,D \rangle = \sum_{i=1}^{n} D_i E_i \leq r \wedge \langle E,E \rangle = \sum_{i=1}^{n} E_i^2 > r \right]$$

It is therefore sufficient to look for r such that

$$\begin{cases} \sum_{i=1}^{n} D_i E_i \leq r \\ \sqrt{\sum_{i=1}^{n} D_i^2} \sqrt{\sum_{i=1}^{n} E_i^2} > r \end{cases}$$

Such an r exists if

$$\sum_{i=1}^{n} D_i E_i < \sqrt{\sum_{i=1}^{n} D_i^2} \sqrt{\sum_{i=1}^{n} E_i^2}$$

or

$$\langle D,E \rangle < \|D\|_2 \|E\|_2$$

i.e. the inequality of Cauchy-Schwarz must be strict. This is so iff

$$E \notin \{\alpha D \| \alpha \in \mathbb{R}\}$$

Let now $E = \alpha D$, $\exists \alpha \in \mathbb{R}$. Since $E \in DS$, $\alpha \in \mathbb{R}^+$. If $\alpha > 1$, take $r = \langle D,E \rangle = \alpha \|D\|_2^2$. Then, since

$$\langle E,E \rangle > \langle D,E \rangle$$

we have that $E \in R(E,r)$. Hence $R(E,r) \in V_\tau(E)$. But $\langle D,E \rangle = r$, hence $D \notin R(E,r)$. If $\alpha < 1$ then we have the above with D and E replaced : $D = (1/\alpha)E$.

(ii) τ is not T_1

Choose any $D,E \neq 0$ in DS with $E = \alpha D$, $\alpha > 1$. For every $V \in V_\tau(D)$ there is an $n \in \mathbb{N}$, $F_i \in DS$, $r_i > 0$ $(i = 1,...,n)$ such that

$$\mathbf{D} \in \bigcap_{i=1}^{n} \mathbf{R}(\mathbf{F_i},\mathbf{r_i}) \subset \mathbf{V}$$

Hence $\langle F_i,D \rangle > r_i$, $\forall i = 1,...,n$ and hence also $\langle F_i,E \rangle = \alpha \langle F_i,E \rangle > r_i$, $\forall i = 1,...,n$. Hence

$$\mathbf{E} \in \bigcap_{i=1}^{n} \mathbf{R}(\mathbf{F_i},\mathbf{r_i}) \subset \mathbf{V}$$

So (DS,τ) is not a T_1-space.

Since $\tau' = \tau''$ are (even T_2-spaces), we have that $\tau \neq \tau''$. □

Note :

In the above proof we had to check carefully if $D \in R(D,r)$. This looks trivial but it is not so. It is even not always true. Indeed $D \in R(D,r)$ iff $\langle D,D \rangle > r$, i.e. iff $\|D\|_2 > \sqrt{r}$. So it can easily be that $R(D,r) \neq \phi$, $R(D,r) \in \tau$ but $D \notin R(D,r)$.

APPENDIX E

PROOF OF THE RESULTS OF EXAMPLE 3.5.2

Theorem 3.9 :

$\forall j = 1,...,n, \forall r > 0$:

$$R(D_j,r) = \left\{ d_i \in DS \mid i \in \{1,...,n\} \cap \left]2jr-j, \frac{j}{2r-1}\right[\right\}$$

forming a subbasis for τ.

Proof :

$\forall i,j \in \{1,...,n\}, \forall r > 0$

$$D_i \in R(D_j,r) = \{D \in DS \mid sim(D,D_j) > r\}$$

$\Leftrightarrow$

$$sim(D_i,D_j) = \frac{i + j}{2 \max(i,j)} > r \qquad (23)$$

a) $i < j$

Then (23) is equivalent to

$$j > i > 2jr - j$$

b) $i > j$

Then, by (a) and reversing i and j, we find

$$i > j > 2ir - i$$

Hence

$$j < i < \frac{j}{2r-1}$$

□

Theorem 3.10 :

Let d be the metric of τ'.

Then, $\forall i,k = 1,...,n$:

$$d(D_i,D_k) = \frac{|i - k|}{2 \max(i,k)}$$

Proof :

$\forall i,k = 1,...,n$, by definition of d,

$$d(D_i,D_k) = \sup_{j=1,...,n} \left| \frac{i + j}{2 \max(i,j)} - \frac{k + j}{2 \max(k,j)} \right|$$

Suppose first that $i \leq k$.

(1) $j \leq i$

Then

$$\sup_{j=1,...,i} \left| \frac{i + j}{2 \max(i,j)} - \frac{k + j}{2 \max(k,j)} \right| = \max_{j=1,...,i} \left| \frac{i + j}{2i} - \frac{k + j}{2k} \right| = \frac{k - i}{2k}$$

(2) $i \leq j \leq k$

Then

$$\sup_{i \leq j \leq k} \left| \frac{i + j}{2 \max(i,j)} - \frac{k + j}{2 \max(k,j)} \right|$$

$$= \max_{i \leq j \leq k} \left| \frac{i + j}{2j} - \frac{k + j}{2k} \right|$$

$$= \max_{i \leq j \leq k} \frac{|ik - j^2|}{2kj} = \frac{k - i}{2k}$$

the maximum being obtained in $j = i$ as well as in $j = k$.

(3) $j \geq k$

$$\sup_{j=k,\ldots,n} \left| \frac{i+j}{2\ \max(i,j)} - \frac{k+j}{2\ \max(k,j)} \right| = \max_{j=k,\ldots,n} \left| \frac{i+j}{2j} - \frac{k+j}{2j} \right| = \frac{k-i}{2k}$$

So, if $i \leq k$,

$$d(D_i,D_k) = \frac{k-i}{2k}$$

For $i \geq k$ we hence have (since $d(D_i,D_k) = d(D_k,D_i)$)

$$d(D_i,D_k) = \frac{i-k}{2i}$$

This proves the theorem. □

Corollary 3.11 :

$\tau = \tau'' = \tau' = \wp(DS)$.

Proof :

It suffices, by theorem 2.5, to show that $\tau = \wp(DS)$. Now $\forall j = 1,\ldots,n$ and

$$r \in \left] \frac{j+\frac{1}{2}}{j+1}, 1 \right[\neq \phi$$

we have that $R(Q,r) = \{Q\}$.

This follows readily from theorem 3.9. □

Note that also τ' is the discrete topology but that d is not the discrete metric.

APPENDIX F

PROOF OF THE RESULTS OF EXAMPLE 3.5.3

Theorem 3.12 :

$\forall j = 1,...,n, \forall r > 0$

$$R(D_j,r) = \left\{ D_i \in DS \mid i \in \{1,...,n\} \cap \left] j - \frac{1}{\tan \frac{\pi r}{2}} , j + \frac{1}{\tan \frac{\pi r}{2}} \right[\right\}$$

which forms a subbasis for τ.

Proof :

$\forall i,j = 1,...,n$

$$D_i \in R(D_j,r) \leftrightarrow \frac{2}{\pi} \operatorname{Arctan} \left(\frac{1}{|i - j|} \right) > r$$

(a) $i < j$

Then

$$D_i \in R(D_j,r) \leftrightarrow \tan \frac{\pi r}{2} < \frac{1}{j - i} \leftrightarrow j - \frac{1}{\tan \frac{\pi r}{2}} < i < j$$

An analogous argument yields the right hand side of the open interval in (24), in case $j > i$. □

Theorem 3.13 :

Let d be the metric of τ'. Then, $\forall i,k = 1,...,n$:

$$d(D_i,D_k) = \frac{2}{\pi} \operatorname{Arctan} |i - k|$$

Proof :

$\forall i,k = 1,\ldots,n$:

$$d(D_i,D_k) = \max_{j=1,\ldots,n} \frac{2}{\pi} \left| \text{Arctan}\left(\frac{1}{|i-j|}\right) - \text{Arctan}\left(\frac{1}{|j-k|}\right) \right|$$

$$= \frac{2}{\pi} \text{Arctan} \max_{j=1,\ldots,n} \tan\left(\left| \text{Arctan}\left(\frac{1}{|i-j|}\right) - \text{Arctan}\left(\frac{1}{|j-k|}\right) \right| \right) \quad ,$$

since Arctan is an increasing function.

By the two lemmas below we have that $\forall i,j,k = 1,\ldots,n$

$$\tan\left(\left| \text{Arctan}\left(\frac{1}{|i-j|}\right) - \text{Arctan}\left(\frac{1}{|j-k|}\right) \right| \right)$$

$$= \left| \tan\left(\text{Arctan}\left(\frac{1}{|i-j|}\right) - \text{Arctan}\left(\frac{1}{|j-k|}\right)\right) \right|$$

$$= \left| \frac{|j-k| - |i-j|}{|i-j|\,|j-k| + 1} \right|$$

using the formula

$$\tan(\alpha - \beta) = \frac{\tan\alpha - \tan\beta}{1 + \tan\alpha \tan\beta}$$

If we split up the cases as in the previous example we find that, in all cases that $i \leq k$ we have that

$$d(D_i,D_k) = \frac{2}{\pi} \text{Arctan}(k - i)$$

The analogue result is true if $i \geq k$, yielding the proof of the theorem. □

Lemma A :

$$|\tan \alpha| = \tan |\alpha| \quad \text{if} \quad \alpha \in \left]-\frac{\pi}{2}, \frac{\pi}{2}\right[\quad .$$

Lemma B :

$$\text{Arctan}\left(\frac{1}{|i-j|}\right) - \text{Arctan}\left(\frac{1}{|j-k|}\right) \in \left]-\frac{\pi}{2}, \frac{\pi}{2}\right[\quad .$$

Proof :

For i,j,k $\in$ {1,...,n},

$$|i-j|,\ |j-k| \in [0,n-1] \quad .$$

From this the conclusion follows easily. □

Corollary 3.14 :

$\tau = \tau'' = \tau = \varrho(DS)$

Proof :

$$R(D_j,r) = \{D_j\} \quad \text{if} \quad r > \frac{1}{2} \quad :$$

indeed, then

$$j - 1 < j - \frac{1}{\tan \frac{\pi r}{2}}$$

$$j + 1 > j + \frac{1}{\tan \frac{\pi r}{2}}$$

Hence $\tau = \varrho(DS)$ and hence also $\tau' = \tau'' = \varrho(DS)$ by theorem 2.5. □