

# DARE: Distance and Angle Retrieval Environment: A Tale of the Two Measures

Jin Zhang\* and Robert R. Korfhage†

School of Information Sciences, University of Pittsburgh, 135 N. Bellefield Avenue, Pittsburgh, PA 15260.  
E-mail: jzhang@sis.pitt.edu

**This article presents a visualization tool for information retrieval. Some retrieval evaluation models are interpreted in the two-dimensional space comprising direction and distance. The two different similarity measures—angle and distance—are displayed in the visual space. A new retrieval means based on the visual retrieval tool, the controlling bar, is developed for a search.**

## Introduction

Consider Figure 1. Which three documents are most similar to the query,  $Q$ ? It depends, of course, on the similarity measure that one uses. However, when this figure has been shown informally at professional conferences, the majority of the audience—up to 90%—chooses documents  $D2$ ,  $D4$ , and  $D6$ , a choice obviously based on distance rather than angle. Why, then, is the cosine measure so dominant in vector retrieval research?

The cosine measure and a distance measure are clearly sensitive to different aspects of documents. Yet a search of the literature on vector retrieval has not revealed a clear argument for the use of the cosine measure rather than a distance measure. In this article we do not propose to make this argument, but rather to present a user interface that incorporates an angular and distance measure, thus allowing the user to judge documents by both measures.

Information retrieval has developed rapidly in recent years. Development is reflected in following aspects: the types and number of information systems have increased dramatically, the ways of information organization have become varied and complicated, and information retrieval evaluation models and means have become not only more powerful but also more sophisticated.

---

\* To whom all correspondence should be addressed. Current address: Jin Zhang, School of Library and Information Science, University of Wisconsin–Milwaukee, 2400 E. Hartford Ave., Milwaukee, WI 53211.

† Deceased.

Received July 17, 1998, revised March 8, 1999, accepted March 8, 1999.

© 1999 John Wiley & Sons, Inc.

However, users hope that the systems are more flexible; they do not care about the complicated mathematical foundations or the calculation of model metrics. Under the circumstances, some of the main trends in the information retrieval field are: (a) how to make systems more user friendly, more transparent to users; and (b) how to provide effective and convenient mechanisms for users to understand the information retrieval process, compare effectiveness of the different evaluation models, dynamically adjust search strategy, and finally achieve satisfactory retrieval results.

To visualize the information retrieval process, display the document distribution in a certain scope, and offer the means to browse the relevant documents are some of the effective ways to achieve the aims. Now, many researchers on the visualization of information retrieval have achieved significant results (Ahlberg & Shneiderman, 1994; Arent et al., 1984; Chalmers & Chitson, 1992; Furnas & Rauch, 1998; Hearst & Karadi, 1997; Heath et al., 1995; Allen, 1998; Benford et al., 1995; Bradley, 1996; Brodbeck et al., 1997; Chalmers, 1993; Hearst, 1995; Helfman, 1994; Koshman, 1996; Nuchprayoon, 1996).

Here, we present a tool for visualization of information retrieval. It provides an interactive visual interface for the users to filter irrelevant documents, view retrieval results, and control the retrieval process on the basis of the distance between document and query as well as a particular angle in document vector space.

So far, there is no visualization model that can simultaneously reflect the two basic measures, distance and angle, which present two distinct methods to judge similarity. *DARE* is the first visualization tool that can provide such functions.

Like other models, the *DARE* model has its inherent weaknesses, ambiguity being one of them. Ambiguous interpretation in the visualization space is a side-effect of the reduction of dimensionality of the document vector space.

## Description and Features of the Tool

Reduction of the dimensionality of the document vector space is one of the underlying parts of visualization. In the

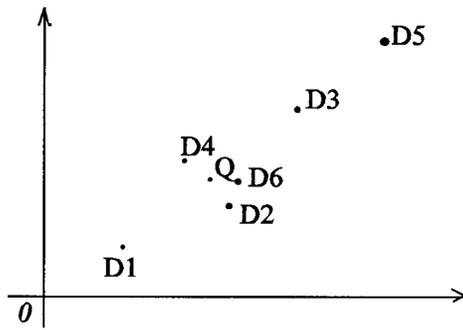


FIG. 1. Judgement of document relevance.

high-dimensional document space, describing the position of a document needs multiple parameters. The problem is that when the dimensionality exceeds 3, it is difficult to visualize the documents in the space. *GUIDO* (Nuchprayoon & Korfhage, 1994) successfully reduces the dimensionality of document space to a comprehensible two-dimensional distance display, the distances being those between a document and either a query or a reference point in the document vector space. *VIBE* (Olsen & Korfhage, 1994), by focusing directly on the similarity ratios of a document to a query and other reference points, can represent any document in relatively low dimensional space. In the model called *BIRD* (Kim & Korfhage, 1994), a more simple and straightforward method is used to reduce the dimensionality, just by picking two meaningful terms to construct a two-dimensional document separator, which is used to filter and select documents.

In our model, the documents are positioned in a two-dimensional plane. It is the visual space. Documents are projected and the retrieval models are interpreted within it. The X-axis of the plane is an angle determined by a document, the query and another special point, a reference point, or the origin of the plane. The Y-axis is the distance from a document to the query.

Before we further delineate the model, it is necessary to introduce several important concepts. The *special view point (SVP)* is a fixed, explicit, and meaningful point in the document vector space, from which documents and reference points are viewed and corresponding directions are measured. The importance, definition, formation, and influences of the reference point on the information retrieval were discussed in detail (Korfhage, 1997).

Selection of the *SVP* would directly influence the relations between the reference points and documents as well as their perceived positions in the visualization space. Two basic modes are considered: origin based, and query based. The former mode is employed when there is only one reference point, the query (generally speaking, query is a special kind of reference point). In this case, the *SVP* is the origin. When two or more reference points exist in retrieval, the query-based mode can be applied. In this case, the *SVP* is the query. This mode is more reasonable, because the query is within a relevant document cluster, and the origin

is relatively far away from it. The final relevancy evaluations of the documents rely on the distances and the directions between the document and reference points or the query rather than the origin. Therefore, the query-based mode better reflects the nature of information retrieval.

The *reference axis (RA)* is another important concept. As we know, any angle in space is formed by two lines that intersect in the space. The *RA* is a special line in the document vector space. It is on the basis of *RA* that the angles of documents in the visualization space are calculated and the final retrieval evaluation is made. The *RA* is determined by two points: one is the *SVP*, which is located at the end of the axis; another is a reference point chosen by user. In the origin-based mode, this reference point is the query.

So in the two-dimensional visualization space, the X-axis value of a document is the angle between the *RA* and the line from the *SVP* to the document itself. Because of the symmetry, the range of the X-axis is from zero to  $\pi$ , which makes the visualization space simpler. Icons in the visualization space represent documents and reference points in the document space. Reference points and documents have different types of icons (see Fig. 2): (a) is the document vector space, (b) is the visualization space.

No matter how high the dimensionality of a document vector space is, the two important parameters of a document, the distance and the angle, are always available.

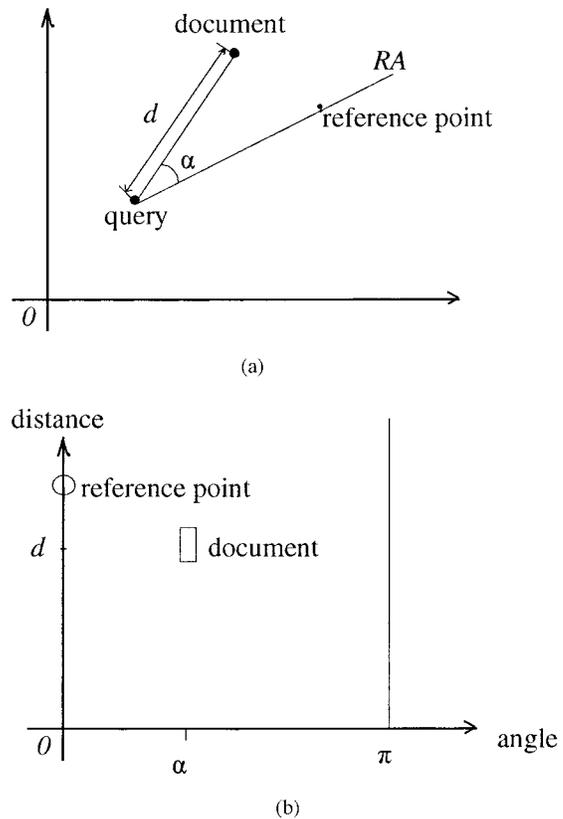


FIG. 2. (a) Document and query in a document space. (b) Document and query in the visual space.

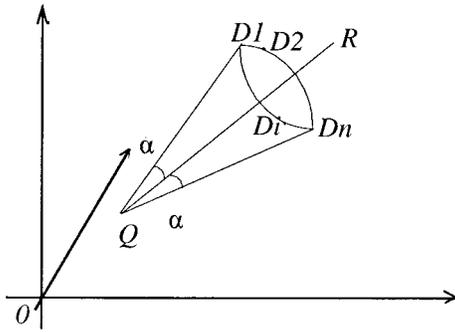


FIG. 3. Ambiguity demonstration.

One problem with any retrieval system is that documents that are the same with respect to the similarity measure or measures used may still differ widely in other respects. Unless additional processing is done, such documents may well be mixed together helter-skelter. For example, in Boolean retrieval, a query for “A or B or C” will mix together documents having one, two, or all three terms, unless further processing is done to separate the subsets out. Even then, the additional processing produces, at best, only a partial ordering of the documents.

A similar situation arises in the combined use of angular and distance measures. In Figure 3, documents  $D_1, D_2, \dots, D_i, \dots, D_n$  are all equivalent with respect to both distance and angle from the query,  $Q$ . Because two similarity measures are being used, there are two methods of partially ordering the documents.

#### Distance Method

A document from an overlaid document set, respectively calculate the distances from this document to the remaining documents in the set, reorganize them by sorting the distances in ascending order, so users can identify and distinguish them.

#### Angle Method

Select a line, for example,  $D_1Q$  in Figure 3, then compute the angles between line  $D_1Q$  and  $D_iQ$ ,  $i = 1, 2, \dots, n$ , where  $n$  is the number of the overlaid documents. Then use the same strategy in the distance method to get a sorted angle list.

Browsing retrieved documents is important and necessary when users make decisions on whether the documents are useful. Although query and reference points are objective criteria to judge whether the documents are relevant or not, the ultimate decision depends on the users, so the visualization model should provide a browsing function for users. The zoom function offers a useful means to “expand” an activated document from its title and keywords to an abstract, even to full text.

Different metrics, such as

$$L_1: \delta(X, Y) = \sum_{i=1}^n |x_i - y_i| \quad (1)$$

$$L_2: \delta(X, Y) = \left\{ \sum_{i=1}^n |x_i - y_i|^2 \right\}^{1/2} \quad (2)$$

$$L_\infty: \delta(X, Y) = \text{MAX}|x_i - y_i| \quad (3)$$

can be used in the document space; whichever is chosen, a given document will have a unique point in the visualization space. However, the position of this point will vary, depending on the metric used. Several different metrics can be displayed simultaneously, as *GUIDO* does, enabling the users to judge the effect of the metrics.

### Visual Display of Information Retrieval Evaluation Models

One of the important tasks of the visualization is to interpret a variety of retrieval evaluation models. We now address this issue.

#### Angle Evaluation Model

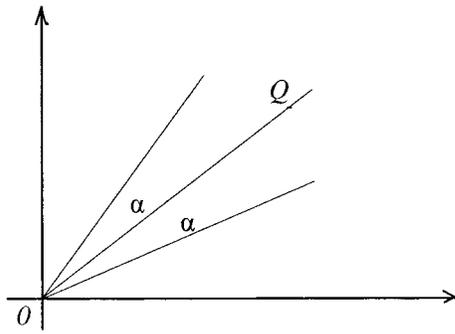
The two methods measure the similarity from different points of view. The angle measure based on direction is not compatible with a distance measure, depending on the length of a document vector space. Documents that are close by one similarity measure may be widely separated by the other.

This model uses both angle and distance as the coordinates of documents and reference points in the display space, enabling comparison of the documents by both measures.

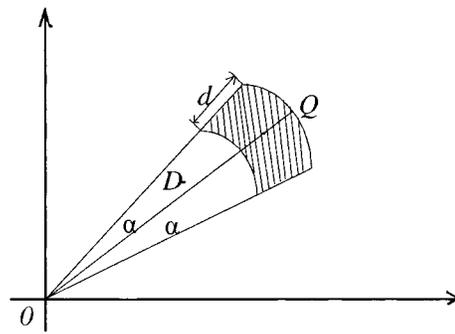
In the cosine evaluation model, documents are retrieved when their angles vis-à-vis the  $QO$  line are smaller than a given threshold  $\alpha$ , where  $Q$  is the query and  $O$  is the origin [see Fig. 4(a)]. The retrieved documents are scattered in the shaded part of Figure 4(b).

Observe that in the traditional cosine model, the angle is the only way to control the evaluation. The impact of distance on the retrieval is ignored, which results in some problems in evaluating the similarity. For example, the cosine is suitable only for comparing similarity between documents of different sizes. Suppose, for example, that a document  $D$  is far from a query  $Q$  in terms of distance, but that the angular separation is small [Fig. 5(a)]. The user’s true need is that any retrieved documents are not only close in angle ( $\leq \gamma$ ), but also in distance ( $\leq d$ ). The traditional cosine model cannot guarantee both, and would erroneously include the  $D$  in the retrieved set.

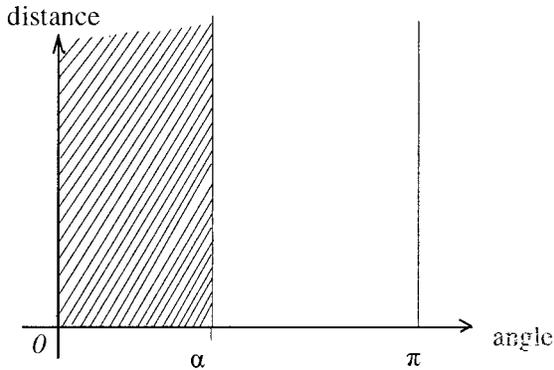
In the model, the user’s need is easily met by controlling both angle and distance variables. The new retrieved documents in the shaded part of Figure 5(a) are scattered in the shaded part of Figure 5(b). The document  $D$  is excluded.



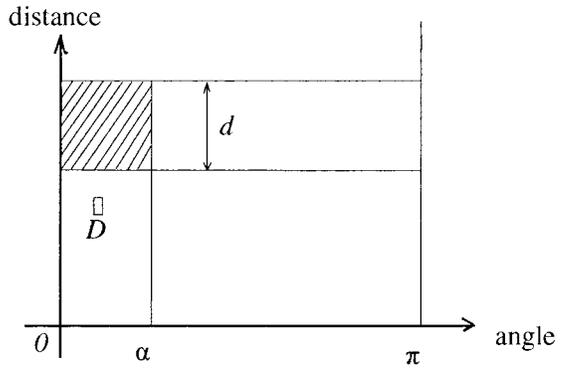
(a)



(a)



(b)



(b)

FIG. 4. (a) Cosine model. (b) Cosine model in the visual space.

FIG. 5. (a) Document *D* in a document space. (b) Document *D* in the visual space.

### Distance Evaluation Model

From Figure 1, we can clearly see documents *D1* and *D2* are very similar in terms of the angle measure, and are quite dissimilar in terms of the distance measure. This suggests that using the different measures could produce quite different results. People sometimes need both measures to meet their needs. Therefore, distance and related measures are necessary for an information retrieval system.

Before discussing the models with added reference points, we delineate what happens when a circle (more generally, a hypersphere) in document vector space is mapped onto the visualization space.

Suppose a condition exists where the distance from a query *Q* to the origin *O* is *h*, the radius under which documents lie within the retrieval space defined by the query is *r*, *P* is any a point on the contour of the circle,  $\alpha$  is the angle between *PO* and *OQ*,  $\beta$  is the angle between *PQ* and *QO*, and *d* is the distance from *P* to *O* (see Fig. 6).

Obviously we have the following relations:

$$r * \sin \beta = d * \sin \alpha \quad (4)$$

$$r * \cos \beta + d * \cos \alpha = h \quad (5)$$

From Equations (4) and (5):

$$r * \cos \beta = h - d * \cos \alpha$$

$$r^2 = h^2 - 2hd * \cos \alpha + d^2$$

$$r^2 - r^2 \sin^2 \beta = h^2 - 2hd * \cos \alpha + d^2 \cos^2 \alpha$$

$$d = h * \cos \alpha \pm (r^2 - h^2 \sin^2 \alpha)^{1/2} \quad (6)$$

The maximum value of *d* is *h* + *r*, and minimum value of *d* is *h* - *r*, when  $\alpha$  is equal to zero.

When  $r^2 = h^2 \sin^2 \alpha$ , *d* has only one solution.

Because  $r^2 - h^2 \sin^2 \alpha$  is always larger than or equal to zero, therefore:

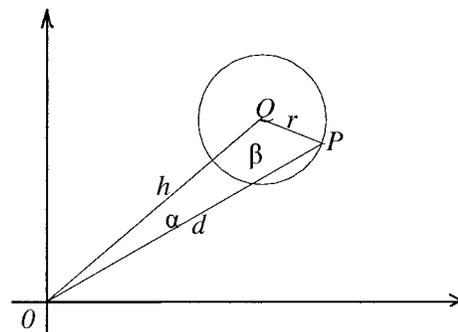


FIG. 6. Distance model.

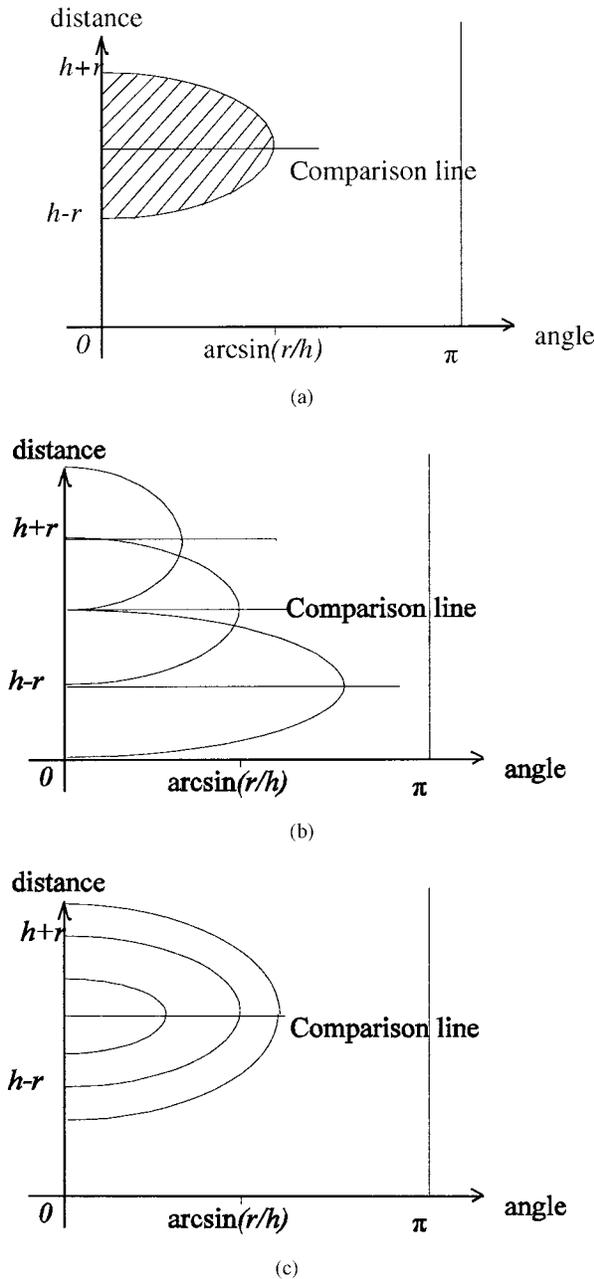


FIG. 7. (a) Distance model in the visual space. (b) Impact of  $h$  on the equivalent curve. (c) Impact of  $r$  on the equivalent curve.

$$r^2 - h^2 \sin^2 \alpha = > 0$$

$$\sin \alpha = < r/h \quad (r, h = > 0) \quad (7)$$

The maximum value of  $\alpha$  is  $\arcsin(r/h)$ .

The shape on which a circle is mapped from document vector space to visualization space is like a "bullet" [see Fig. 7(a)].

The length of the "bullet" is  $\arcsin(r/h)$ ; the width is  $2r$ .

Let us discuss the following two cases in detail: As  $r$  is fixed and  $h$  increases, the length of the corresponding "bul-

let" becomes smaller, and the width of the "bullet" remains the same ( $2r$ ), and the comparison line ( $y = h$ ) of the "bullet" moves up [see Fig. 7(b)].

This case suggests that a circle in the document space moves away from the origin, but the radius of the circle is a constant.

When  $h$  is fixed and  $r$  increases, both the length and width of the corresponding "bullet" become larger while the comparison line of the "bullet" is unchanged ( $y = h$ ) [see Fig. 7(c)].

This case suggests that the distance between a circle and the origin remains the same, but its area increases or decreases, depending on whether  $r$  is held constant.

### Conjunction and Disjunction Evaluation Models

Now we address how the traditional conjunction and disjunction models are visualized in this proposed model.

Conjunction evaluation model (or intersection model):

$$(\delta(D, R1) \leq k) \vee (\delta(D, R2) \leq k) \quad (8)$$

or:

$$\min(\delta(D, R1), \delta(D, R2)) \leq k$$

$R1$  and  $R2$  are reference points, and  $k$  is the threshold of retrieval (see Fig. 8).  $\delta(X, Y)$  stands for the distance between  $X$  and  $Y$  in the document space. Because we shall visualize two circles in the visualization space, in most cases the symmetry between  $r$  and  $h$  no longer exists. In other words, if we choose  $OR1$  as the  $RA$ , for example, the circle whose center is  $R2$  is asymmetrical vis-à-vis the  $RA$ ,  $OR1$ . Therefore, the range of the  $X$ -axis in the visualization space should be from  $-\pi$  to  $\pi$  so that the circle can be accurately displayed in the visualization space. The result of Equation (8) is illustrated in Figure 9(a).

Disjunction evaluation model (or union model):

$$(\delta(D, R1) \leq k) \wedge (\delta(D, R2) \leq k) \quad (9)$$

or:

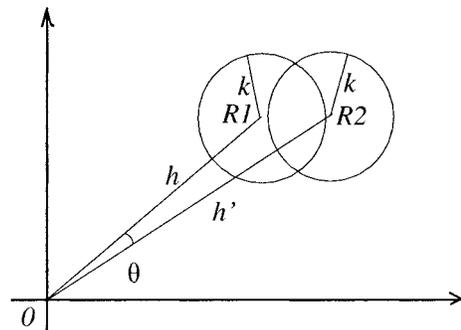


FIG. 8. Conjunction and disjunction models.

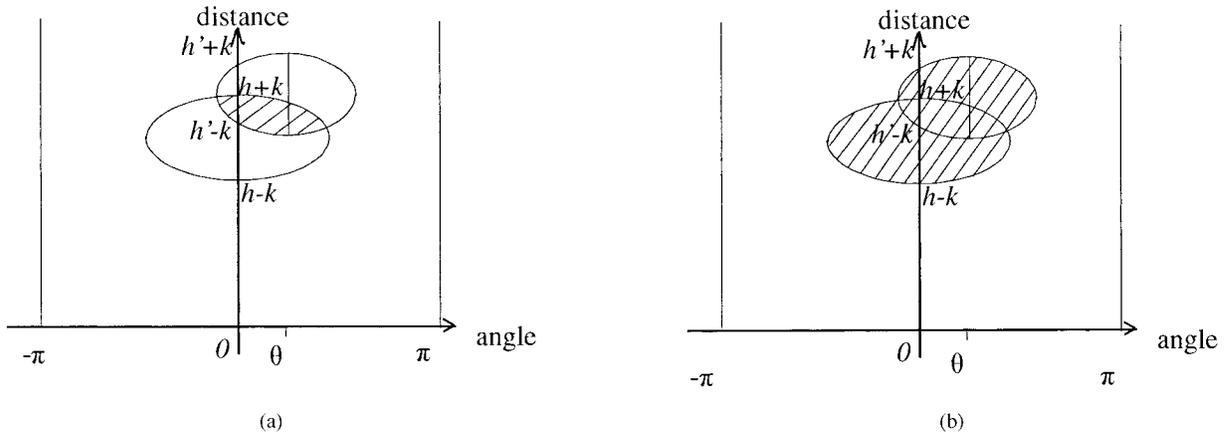


FIG. 9. (a) Connunction model in the visual space. (b) Disjunction model in the visual space.

$$\max(\delta(D, R1), \delta(D, R2)) \leq k$$

The result of Equation (9) is illustrated in Figure 9(b).

Although  $k$  is really the radius of each circle, the different positions and shapes in Figure 9 are due to the different  $h$ ,  $\arcsin(k/h)$  in each circle, and the angle  $\theta$ .

#### Elliptical Evaluation Model

In the elliptical model, two reference points  $R1$ ,  $R2$  are required. We use  $d1$ ,  $d2$  as the distances from a point  $P$  on the ellipse to the reference points  $R1$  and  $R2$ , and  $\alpha 1$  and  $\alpha 2$  as the angles between the  $PR1$ ,  $R1R2$  lines and  $R2R1$  lines, respectively (see Fig. 10).

According to the definition of the ellipse we have:

$$d1 + d2 = k, \quad (10)$$

where  $k$  is the threshold. We also have the following equality relations:

$$d1 * \sin \alpha 1 = d2 * \sin \alpha 2, \quad (11)$$

$$d1 * \cos \alpha 1 + d2 * \cos \alpha 2 = a \quad (12)$$

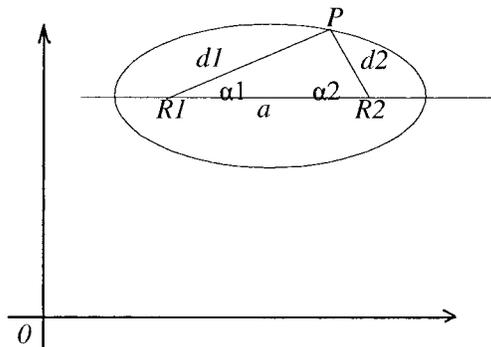


FIG. 10. Ellipse model.

where  $a$  is the distance from  $R1$  to  $R2$ , constant for a given ellipse.

From Equations (11) and (12):

$$d2^2 = a^2 - 2ad1 * \cos \alpha 1 + d1^2 \quad (13)$$

And from Equations (13) and (10):

$$a^2 - 2ad1 * \cos \alpha 1 = k(d2 - d1) \quad (14)$$

Thus, from Equations (14) and (10):

$$\cos \alpha 1 = \frac{kd1 - kd2 + a^2}{2ad1}$$

$$\cos \alpha 1 = \frac{(a^2 - k^2)}{2a} * \frac{1}{d1} + \frac{k}{a} \quad (15)$$

Function (15) is continuous.

When  $\alpha 1 = 0$ , then  $d1 = (k + a)/2$ .

When  $\alpha 1 = \pi/2$ , then  $d1 = (k^2 - a^2)/(2k)$ .

When  $\alpha 1 = \pi$ , then  $d1 = (k - a)/2$ .

In the construction of the visualization space, using the query-based mode,  $R1$  is the  $SVP$ ,  $R1R2$  is the  $RA$ , and  $(\alpha 1, d1)$  is the coordinate of  $P$  in the visualization space. The boundary area defined by the ellipse is then mapped on the shaded part of Figure 11.

In Figure 11:

$$M1 = \left( 0, \frac{k + a}{2} \right); \quad M2 = \left( \pi/2, \frac{k^2 - a^2}{2k} \right);$$

$$M3 = \left( \pi, \frac{k - a}{2} \right).$$

#### A New Information Retrieval Evaluation Model

Notice that there is a very interesting phenomenon: a straight line in the visualization space has a meaningful

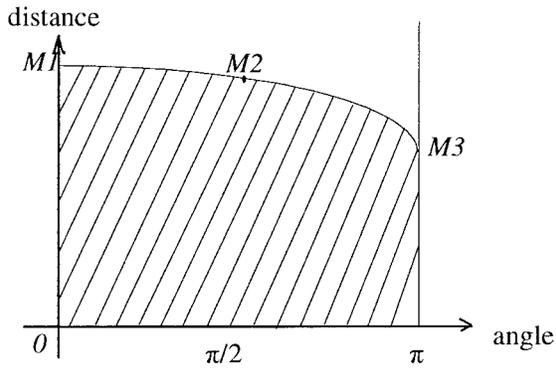


FIG. 11. Ellipse model in the visual space.

interpretation in document vector space. It can yield a contour, which is similar to an oval, but is not an oval. The reference axis is its central line; the maximum length and width of the contour depend on the slope of the line in the visualization space. We call the line A “controlling bar” [see Fig. 12(a) and (b)].

This suggests that by controlling the position and the slope of the controlling bar in the visualization space, users can effectively increase or shrink the boundary area around the two reference points in a document vector space at will.

Let us discuss this control in detail. Suppose we have a straight line:

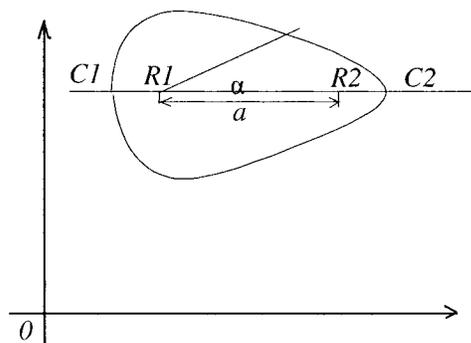
$$d = b\alpha + c \quad (16)$$

The  $d$  is a distance variable,  $\alpha$  is an angle variable, the  $b$  and  $c$  are constants.  $R1, R2$  are two reference points,  $R1R2$  is the  $RA$ ,  $R1$  is the  $SVP$ .

In a document vector space the contour intersects with the  $RA$  in two points:  $C1(\pi, d1)$  and  $C2(0, d2)$ .

$$d1 = \delta(R1, C1) \quad (17)$$

$$d2 = \delta(R1, C2) \quad (18)$$



(a)

In the visualization space, the positions of  $C2$  and  $C1$ , which affect the shape of the contour, are determined by the intersections of the line  $d = b\alpha + c$  with the lines  $\alpha = 0$  and  $\alpha = \pi$ , respectively. So we have:

$$c = \delta(R1, C2) \quad (19)$$

$$\delta(R1, C1) = b\pi + c \quad (20)$$

If a user considers that one of the reference points is more important than the other in retrieval, he or she can extend the corresponding part of the contour by moving  $C1$  or  $C2$  up along the axis:  $\alpha = \pi$  or  $\alpha = 0$ ; or vice versa. But the minimum values of  $C1$  and  $C2$  are respectively limited to  $I1(\pi, 0)$  and  $I2(0, a)$ ; here  $a$  is the length of  $R1R2$  [see Fig. 11(a) and (b)].

When  $b = 0$ , then  $d = c$ .

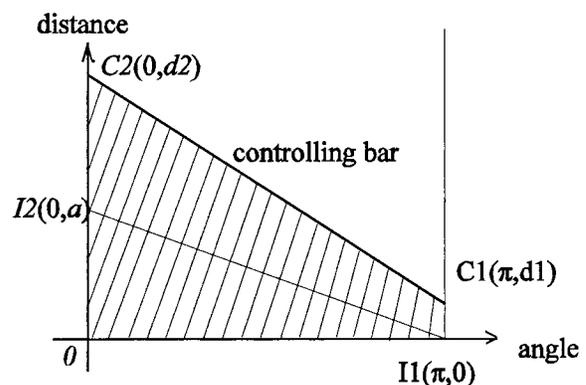
When  $d = c$  there is a horizontal line in the visualization space; in that case, the corresponding contour in the document vector space becomes a circle, whose center is  $R1$  and the radius is  $c$ .

We now address another important issue: under what conditions a document is retrieved in the document space when the controlling bar is manipulated by users.

In the document space, suppose we have a document  $D(x_{11}, x_{12}, x_{13}, \dots, x_{1n})$ , and two reference points  $R1(x_{21}, x_{22}, x_{23}, \dots, x_{2n})$ , and  $R2(x_{31}, x_{32}, x_{33}, \dots, x_{3n})$ , where  $n$  is the dimensionality of a document vector space,  $\beta$  is the angle formed by the  $DR1$  and  $R1R2$ , we have:

$$\cos \beta = \frac{\sum_{i=1}^n \{[x_{1i} - x_{2i}][x_{3i} - x_{2i}]\}}{[\sum_{i=1}^n (x_{1i} - x_{2i})^2]^{1/2} [\sum_{i=1}^n (x_{3i} - x_{2i})^2]^{1/2}}$$

$$\beta = \text{Arccos} \frac{\sum_{i=1}^n \{[x_{1i} - x_{2i}][x_{3i} - x_{2i}]\}}{[\sum_{i=1}^n (x_{1i} - x_{2i})^2]^{1/2} [\sum_{i=1}^n (x_{3i} - x_{2i})^2]^{1/2}} \quad (21)$$



(b)

FIG. 12. (a) Controlling bar model. (b) controlling bar model in the visual space.

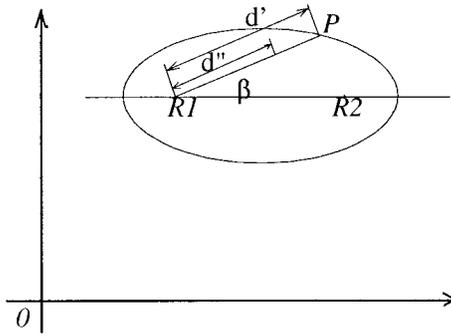


FIG. 13. Threshold control of control bar model.

If a point  $P$  on the contour has this angle,  $\beta$ , vis-à-vis the RA,  $R1R2$ , the distance  $d'$  from  $P$  to the  $SVP$ ,  $R1$ , should be:

$$d' = b\beta + c$$

$$d' = b \operatorname{Arccos} \frac{\sum_{i=1}^n \{[x_{1i} - x_{2i}][x_{3i} - x_{2i}]\}}{[\sum_{i=1}^n (x_{1i} - x_{2i})^2]^{1/2} [\sum_{i=1}^n (x_{3i} - x_{2i})^2]^{1/2}} + c \quad (22)$$

In fact, the distance between the document  $D$  and the  $SVP$ ,  $R1$ , is:

$$d'' = \delta(D, R1) \quad (23)$$

$$d'' = \left[ \sum_{i=1}^n (x_{1i} - x_{2i})^2 \right]^{1/2}$$

From Equations (22) and (23), we can conclude that when  $d' \leq d'' = \delta(D, R1)$ , the document  $D$  shall be retrieved, i.e., it is within the contour determined by the controlling bar in the visualization space (see Fig. 13).

In summary, the bigger the slope of the controlling bar, the fatter and larger the corresponding side of the contour. When  $d' \leq d''$  the document is retrieved.

## Conclusion

To facilitate user's retrieval, visualization of information retrieval is one of the most important means. The visualization model presented here can display both the distance and angle (cosine) measures. In addition, it improves the retrieval effectiveness of the cosine model when the impacts of both direction and distance of a document on retrieval are taken into account. It also can visualize some popular information retrieval evaluation models such as the conjunction, disjunction, elliptical models, etc. Furthermore, it provides a new retrieval means. It is convenient for users to manipulate the controlling bar to adjust the scope and the emphasis of the

retrieval. Some evaluation models such as the *Cassini* model can not be displayed effectively in this visualization model. The tool allows users to browse, and to zoom in on the retrieved documents.

Users can use multiple reference points in retrieval in the *DARE* visual environment. There are two basic strategies to deal with them. The first is to take each new reference point along with the query to yield a new *RA*, and reconstruct the visualization space. The second is to use three reference points (including the query) to construct a plane called the reference plane. The coordinates of a document in the visualization space are obtained in the following way: the  $Y$ -axis value is still the distance from the document to the query, but the  $X$ -axis value is the angle measured between the reference plane and the line from the query to the document (note that query is on the plane). However, the penalty of this method is the computational complexity.

Further research on the model will include the coordination strategy of the visual display, ambiguity of the display, and the new retrieval method based on the tool.

## References

- Ahlberg, C., & Shneiderman, B. (1994). Visual information seeking: Tight coupling of dynamic query filters with Starfield displays. In *Proceedings of CHI'94 Human Factors in Computing Systems* (pp. 313–317), Boston: ACM.
- Allen, B. (1998). Information space representation in interactive systems: Relationship to spatial abilities. In *Proceedings of the Third ACM Conference on Digital Libraries* (pp. 1–10), Pittsburgh: ACM.
- Arent, H.C., & Bogaerts, W.F.L. (1984). Concept-based retrieval of hypermedia information: From term indexing to semantic hyperindexing. *Information Processing & Management*, 29(3), 373–386.
- Benford, S., et al. (1995). VR-VIBE: A virtual environment for cooperative information retrieval. In *Eurographics 95* (pp. 123–134).
- Bradley, J.L. (1996). A software architecture for automated information visualization. Ph.D. Dissertation, University of Texas at Austin.
- Brodbeck, D., et al. (1997). Domesticating bead: Adapting an information visualization system to a financial institution. In *Proceedings of IEEE Symposium on Information Visualization* (pp. 73–80), Phoenix, AZ: IEEE.
- Chalmers, M. (1993). Using a landscape metaphor to represent a corpus of documents. In A.U. Frank, & C.I. Canorri (Eds.), *Spatial information theory: A theoretical basis for GIS* (pp. 377–390), Berlin: Springer Verlag.
- Chalmers, M., & Chitson, P. (1992). Bead: Explorations in information visualization. In *Proceedings of the Fifteenth Annual International ACM SIGIR Conference on Research and Development in Information Retrieval* (pp. 330–337), Copenhagen, Denmark: ACM.
- Furnas, G.W., & Rauch, S.J. (1998). Considerations for information environments and the NaviQue workspace. In *Proceedings of the Third ACM Conference on Digital Libraries* (pp. 79–88), Pittsburgh: ACM.
- Hearst, M.A. (1995). TileBars: Visualization of term distribution information in full text information access. In *Proceedings of CHI'95 Human Factors in Computing Systems* (pp. 59–66), Denver, CO: ACM.
- Hearst, M.A., & Karadi, C. (1997). Cat-a-cone: An interactive interface for specifying searches and viewing retrieval results Using a large category hierarchy. In *Proceedings of the 20th Annual International ACM SIGIR Conference on Research and Development in Information Retrieval* (pp. 246–255), Philadelphia, PA: ACM.
- Heath, L., et al. (1995). Envision: A user-centered data base of computer science literature. *Communications of the ACM*, 38(4), 52–53.

- Helfman, J.I. (1994). Similarity patterns in language. In Proceedings 1994 IEEE symposium on visual languages (pp. 173–175), St. Louis, MO: IEEE.
- Kim, H., & Korfhage, R.R. (1994). BIRD: Browsing interface for the retrieval of documents. In Proceedings 1994 IEEE symposium on visual languages (pp. 176–177), St. Louis, MO: IEEE.
- Korfhage, R. (1997). Information storage and retrieval, New York: Wiley Computer Pub.
- Koshman, S. (1996). User testing of a prototype visualization-based information retrieval system. Ph.D. Dissertation, School of Information Sciences, University of Pittsburgh.
- Nuchprayoon, A. & Korfhage, R.R. (1994). GUIDO, a visual tool for retrieving documents. In Proceedings 1994 IEEE computer society workshop on visual languages (pp. 64–71), St. Louis, MO: IEEE.
- Nuchprayoon, A. (1996). GUIDO: A usability study of its basic information retrieval operations. Ph.D. Dissertation, School of Information Sciences, University of Pittsburgh.
- Olsen, K.A., & Korfhage, R.R. (1994). Desktop visualization. In Proceedings 1994 IEEE symposium on visual languages (pp. 239–244), St. Louis, MO: IEEE.