



PROJECT MUSE®

---

Precision and Recall: An Ontological Perspective /  
Précision et rappel: un point de vue ontologique

William Buck

Canadian Journal of Information and Library Science, Volume 41,  
Numbers 1-2, March-June/mars-juin 2017, pp. 42-51 (Article)

Published by University of Toronto Press



➔ For additional information about this article

<https://muse.jhu.edu/article/666448>

# Precision and Recall: An Ontological Perspective

# Précision et rappel : un point de vue ontologique

William Buck  
Talking Book Program (Retired)  
Texas State Library and Archives  
williambuck@my.unt.edu

*Dedicated to Edwin B. Allaire / Dédié à Edwin B. Allaire*

**Abstract:** There is a traditional narrative within information studies regarding precision and recall measures. Precision and recall have been the most commonly used retrieval metrics and are the basis for more complicated and accurate information retrieval evaluations. Relevance, which is the criterion by which both recall and precision are judged, is subject to user interpretation and is context dependent. Although the determination of precision is straightforward, important ambiguities are involved when considering recall. Search evaluation metrics can be parsed into structural components. The interaction of the component parts can be clarified by the application of ontological distinctions. Possibly relevant items not retrieved are most usefully viewed as conceptually dependent parts. Positioned as the denominator of the recall measure, these parts function analogically and are supplemental in character. Although difficult to accurately determine, the recall denominator performs a useful role in the assessment of indexed data structures and other collection types.

**Keywords:** relevance, precision, recall, ontology, part-whole

**Résumé :** Il existe un narratif traditionnel au sein des sciences de l'information concernant les mesures de précision et de rappel. La précision et le rappel sont les mesures le plus couramment utilisées en recherche d'information, et elles constituent également la base d'autres méthodes d'évaluation de plus complexes et plus précises. La pertinence, critère selon lequel sont jugés à la fois le rappel et la précision, est sujette à l'interprétation de l'utilisateur et dépend du contexte. Bien que le calcul de la précision soit relativement simple, celui du rappel comporte des ambiguïtés importantes. Les mesures d'évaluation de la recherche d'information peuvent être segmentées en divers composantes structurales. L'interaction entre ces composantes peut être clarifiée par l'application de distinctions ontologiques. Éventuellement les documents potentiellement pertinents qui n'ont pas été repérés sont considérés comme des éléments conceptuellement dépendants. En tant que dénominateur de la mesure de rappel, ces éléments sont complémentaires et fonctionnent de manière

analogique. Bien qu'il soit difficile de le mesurer avec précision, le rappel joue un rôle important dans l'évaluation des structures de données indexées et autres types de collection.

Mots-clés : pertinence, précision, rappel, ontologie, relation partitive

### Introduction: precision and recall

There is a traditional narrative within information retrieval studies that is familiar and oft repeated. People have information needs of various sorts relating to their lives and interests. Indexes establish searchable structures of concepts (the semantic aspect of the terms) that are contained in information items. Once an information need is translated into a query, the search statement can be addressed to these searchable data structures, such as a library catalogue or a database. The retrieval process begins at this point. The goal of information retrieval research for the most part has been to locate an isomorphic or atomic match between the representation of a query (sometimes called a “compromised” or “expanded” query) and a representation of a document that has been indexed into the retrieval system. Performance measures are then applied to determine how relevant the results are to a patron's interest.

In this parable, relevance becomes the standard bearer of value. A retrieved item is relevant if it contributes to the resolution of an information need. The question then naturally arises as to what could be the best measures for evaluating the procedure. Information retrieval metrics have almost always relied upon some variant of the recall and precision ratios (Baeza-Yates and Ribeiro-Neto 2011). Developed in the 1960s by Cyril Cleverdon at the Cranfield Institute in the United Kingdom, these formulae laid the groundwork for most of the subsequent research on systems evaluation. The concepts in their most basic form can be described in the following way. To determine precision in any given search, the number of relevant items retrieved is divided by the number of all of the items retrieved. This is the precision ratio, which is here expressed as follows: let  $P$  stand for precision,  $L$  for relevance,  $T$  for sum total, and  $i$  for items, so  $P = Li/Ti \times 100\%$ , where the numerator is the number of relevant items retrieved and the denominator is the number of total items retrieved. To determine recall in any given search, the number of relevant items retrieved is divided by the number of all possibly relevant items that were not retrieved. This is the recall ratio, which is here expressed as follows: let  $R$  stand for recall,  $L$  for relevance,  $P$  for possible, and  $i$  for items, so  $R = Li/Pi \times 100\%$ , where the numerator is the number of relevant items retrieved and the denominator is the total number of possibly relevant items in the collection or data structure. The parameters of the method become quickly apparent: if recall is increased, more non-relevant items appear. If precision is increased, more possibly relevant items might be missed. The ordered pairs are inverse ratios (Rowley and Hartford 2008, 294–96).

## Relevance

Precision is not difficult to determine, as one need only compare what the searcher considers to be the relevant retrieved items with the non-relevant retrieved items. Therefore, if among 50 retrieved items a searcher marks as relevant 40, 40 divided by 50 is 0.8, multiplied by 100 yields an 80% precision ratio. The only serious limitation occurs if the searcher quits the review process before viewing enough returned items. In that context, Web search engine precision evaluations can help determine the weighting of terms to obtain a high concentration of relevant items at the top of a ranking (Gómez and Abasolo 2003, 130–31). This is so because precision highlights those results that contain information that searchers quickly recognize as contributing to their interests. Consider the following line of reasoning:

Amanda wants articles on strawberries.

Julie wants articles on fragaria ananassa.

*Amanda and Julie want the same thing.*

The inference is valid. The semantics of ordinary language are adequately represented by the precision measures as they are understood by most searchers of indexed data structures. They are good indicators of the number of keywords that are relevant in an expanded or compromised query and are useful in identifying the synonymy between terms and concepts. Specialized techniques such as relevance feedback, vector similarity measures, and ranking algorithms can be employed to adjust and improve the precision ratio.

Determining recall, by contrast, is more problematic. A troubling question presents itself to the information retrieval researcher—namely, how to determine the recall denominator—since it represents a quantity of unknown items. How can items that are relevant but not returned be known? At this point in the narrative, an appeal is often made for the insertion of “experts” to provide a solution. It is sometimes claimed that this is where the costs of implementing an information system are greatest and that developing a new system or integrating and updating a legacy system is too complex for librarians or library staff, requiring, instead, the expertise of expensive consultants. The ability to evaluate recall would in fact require a comprehensive knowledge of the collection or data structure under consideration. Even if this requirement is met, it is difficult to see how one is not second-guessing both the patron and the system in regard to what is considered relevant.

The concept of relevancy is used as a success criterion. However, determining the relevancy of an item is not a binary truth condition, such that it is always either true or false that a particular item is relevant. Relevancy is subject to user interpretation, and there is typically a continuous function of lesser to greater degrees of relevancy between the two extremes of being completely useless to a searcher and being exactly what the searcher was looking for. Kowalski (2011, 254–56) points out that relevance is always context dependent, and he lists five attributes of relevancy:

- subjective—depends upon a specific user's judgment;
- situational—relates to a user's requirements;
- cognitive—depends on human perception and behaviour;
- temporal—changes over time; and
- measurable—observable at points in time.

Different individuals have differing perspectives on what, to them at a particular time, is helpful information. Even information that from a systems' perspective is directly relevant to a topic may not be relevant to a user who is already familiar with it.

Given the subjective and contextual nature of relevance judgements, it does not seem possible to accurately determine the number of possibly relevant items. On this point, consultants may sometimes demand higher costs from a library or other information centre to continue work on the system. A question that can be asked here is: are the possible items in fact actual? The total number of relevant items in the collection would be the denominator in the recall measure; however, something would have to determine their relevance. The system apparently cannot make this determination, or it would have retrieved them.<sup>1</sup> If one predetermines which items are to be relevant—for example, by placing them in a data set that will be used to judge how relevant a set of retrieved items are—then one is clearly reasoning in a circle. *Eo ipso*, if the system recognized the relevance of an item, it would retrieve it. The success criteria for the recall function determine whether it can return the most relevant results from those items that are possibly relevant, but what is possibly relevant has to be defined in advance of the search. This type of inference—where what is to be shown or proven (in this case, the relevance of items not retrieved) is in some sense predetermined or defined in advance of a conclusion—is considered to be a logical fallacy of the type *petitio principii* (Hurley 2014, 142–44).

The attraction of these concepts is due in large part to the explanatory power that they seem to promise. The ideal information retrieval system would return very high or complete precision for any level of recall. The retrieval set would only contain relevant items, and all of the relevant items available, while passing over the irrelevant items in silence. However, studies continue to show that these measures only work effectively in small collections that are carefully controlled, as opposed to large collections that are continually receiving new material (Kowalski 2011). In addition, the measures were developed in a pre-Internet era when results were produced in batch mode, as opposed to the interactive nature of Web searching today. Current information retrieval techniques using Web analytics have diverged significantly from older search models (Jansen 2009). Given these difficulties, one might wonder why precision and recall wield the fascination that they do for researchers and information studies. I believe that the reason is in large part due to an implicit relationship in the recall function that tantalizes the understanding without becoming explicit. To this relationship I shall now turn.

### Ontology, Part 1: Existence

The determination of what exists and how existents relate to each other is the fundamental task of ontology. An ontological analysis provides a clarification of the nature of the relationships (if any) between entities. However, ontologies as they are used within information systems have important differences from the way ontology is understood in the tradition of Western metaphysics. Ontologies as they are used by information studies are inventory vocabularies for different areas of knowledge. These representation vocabularies are sets of terms that describe the different objects and their relations in a particular domain (Chandrasekaran, Josephson, and Benjamins 1999). One common example is the field of health informatics, where medical terms are grouped according to subject. This use of ontology presupposes the existence of facts that can be enumerated and sorted by various criteria. Understood in this sense, they are content theories, and a content theory must, by definition, have a content. Ontology, as it is used in metaphysics, by contrast, is a determination about what kinds of being exist, without presupposing the existence of anything in particular. One may notice here that ontology as it is used in the first sense will have an inbuilt bias toward content-rich domains because the task is to name the content items. A taxonomy cannot be modelled around an empty class. To avoid this preoccupation, ontology is here understood in a broader metaphysical sense.

Considering the precision and recall ratio, one may ask in this context the following question: what is the nature of the relationship between actually retrieved items and possibly relevant items not retrieved? To explicate this question, I make use of the part-whole relationship that is considered by some researchers to be the closest available candidate among human cultures for a universally recognized relationship (Raybeck and Herrmann 1990). Empirical studies have shown that this relationship or near equivalent is present in some form in all human languages (Wierzbicka 1994, 488–90). Another way of saying this in the context of information theory is that parts and wholes, like space and time, are terms that apply to all domains. The part-whole relation has proven to be a useful and powerful intellectual tool and has been recognized as a fundamental ontological relation since the time of the ancient Greeks. If this relation is rejected, then one has rejected the possibility of an ontological perspective.

Entities are the basic building blocks of ontologies. Although basic, they have structure and can be divided into parts. To meaningfully say that an entity has parts, it must have more than one part. Objects and events have parts, as do abstract entities such as ideas and numbers. Items in a database, the grains in a sand dune, the fractions of a number, and the petals on a flower can be formalized into paronomies. Part structures change over time, raising puzzling questions about identity. Distinguishing the various parts of a composite is not always an easy or obvious task. Physical entities are not the only existents for which a paronomy can be developed. Concepts can also be separated into constituent elements, when, for example, one is considering a set of statements and their semantic referents. Ontology searches for the elements, constituents,

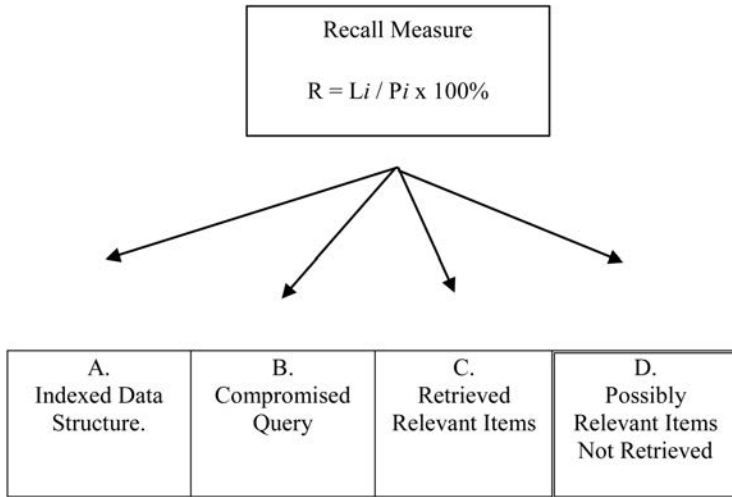


Figure 1: A concept partonomy of the recall measure that shows member collection (C, D) and features of the search activity (A, B). D is shown segmented to illustrate its conceptually dependent status.

or parts of structures. Of these parts, some have a unique or special role to play in forming the structure itself.

There is a special class of parts that is sometimes referred to as “dependent parts” (Pribbenow 1999). These parts are usually non-material, such as the interior of a cup or the keyhole in a door lock. They are in one sense a part like other parts, sometimes even providing the critical function of the object, as in the case with a keyhole or an auditorium space. However, certain material predicates such as colour or weight cannot be attributed to them, but others such as shape sometimes can be. The defining aspect of this class of parts is their dependence on the host object(s). If the host object such as a cup is disassembled, the interior of the cup will not be one of the disassembled parts. Similar issues can be raised about the shadows that objects cast (Casati and Varzi 1997). The ontological status of such entities is conceptual rather than material; however, they perform real functions for the objects of which they are a part.

A set of retrieved relevant items and a set of retrieved non-relevant items can be considered a complex entity, such as a state of affairs. They contain objects (the items) and properties (the relevance or non-relevance of the items); likewise, a set of retrieved relevant items and the set of possibly relevant items that were not retrieved. One may say that the possibly relevant items are a part of the recall measure. Similar types of sentences include “the conclusion is part of the argument” and “the domain is part of the model” (Winston, Chaffin, and Herrmann 1987). Moreover, the concept of a recall measure can be understood as a whole consisting of recognizable parts. The diagram in figure 1 displays this relationship.

In their discussion on meronymic relationships, Winston, Chaffin, and Herrmann (1987) provide a taxonomy of six types of part-whole relations. The recall measure could fit two of them, the member/collection and the feature/activity. In the first case, one criterion offered for membership in a collection is the spatial proximity of the items. Both the retrieved relevant items and the possibly relevant items not retrieved (C and D) are members of the same indexed data structure. In the second case, the measure represents a search activity, and two features of this activity are depicted (A and B). The four listed items have clear relations to each other and are interrelated by function. One of the items, however, has a non-determinate quality.

Possibly relevant items have meaning because in this context existence is a prerequisite for not having properties as well as for having them. They play the same role as other dependent parts that have specific linguistic references. Possible items, even if they are not actual, have the ontological status of existence if only in the sense that they exist for thought. They are not arbitrary, however, as they must conform to the agreed upon definition of relevance. Although there is any number of items that could be possibly relevant, there are items that, depending on the definition of relevance, are not. To speak analogically, there may be an indefinite number of ways to slice an apricot, but none of them produce strawberries.

### **Ontology, Part 2: Independence**

Notions of dependence and independence in ontology can be understood in the following way. Items are independent of each other if they can exist separately from each other (Fine 1995). For an item that has parts, the parts that can exist separately from the remaining parts are said to be independent. A common example is the case of automotive drum brakes. The springs, pins, shoes, and adjusting mechanisms can all be laid out separately for inspection. By contrast, an example commonly given of two parts that are distinct, yet dependent, are the colour and the shape of an object. One never encounters a colour that is not shaped in some way (Lampert 1995, 79–80). Of the four items listed in figure 1, Items A, B, and C can be conceived as existing independently of each other. A query can be composed without addressing it to an indexed data structure. An indexed data structure could be available without a query being submitted to it. A set of retrieved relevant items could be kept in a file apart from either a query or a database. A set of possibly relevant items not retrieved does not have this type of independence. They only appear, as it were, when the recall measure is being considered.<sup>2</sup>

Rescher (1979) has suggested that possibilities attain what reality they have due to their construction by verbal descriptions (in this case, establishing the definition of relevance) and the postulation of their existence by a thinking subject. The possible items are dependent upon being conceptualized. Note that, in this regard, actual items retrieved can be thought or conceptualized without reference to possibly relevant items not retrieved, but not conversely. Whenever one considers the possibly relevant items, there is always an inbuilt reference to the actual ones. Therefore, it would seem that there is no literal



sense in which the possible items are parts of the complex entity that is the recall measure. The whole-part relation is analogical since the possible items are not distinguishable one from the other while they are still unknown. When they become known, they are distinguishable, but then they are no longer merely possible. In one sense, the space that the relevant possibles occupy is not the indexed data structure; rather, they are conceptual referents of the recall measure itself.

Although the possible items are dependent upon their function in the recall measure, they are not useless fictions. If a patient executed a query for which certain items held the cure, and these items were not retrieved, these unfound items would retain their value for the health of the searcher. The possible items exist conceptually; they may also exist as actual items—items that will make a real difference. Consider the following line of reasoning:

Amanda is searching for articles on apricots.  
 Julie is searching for articles on strawberries.  
*Amanda and Julie are searching for articles on something.*

The inference is invalid. “Something” can only validly quantify over matching content. The recall denominator serves as a regulative principle for items that may be relevant but that were missed because of a lack of topical matching in a way that categorical logic cannot account for. Since the measure purports to show a failure of reference, it does not literally refer to anything actual or determinate in the information retrieval system. From this point of view, it is not the case that the recall measure presupposes what is yet to be discovered. Instead, it performs the function of a placeholder for items that were not retrieved due to their analogical or figurative relationship to a query topic.

Possibly relevant items not retrieved are postulated as those items that, once their status is determined to be relevant, could supplement the returned results. Implicit in this idea is that a set of returned results can never be completely known to exhaust all of the possibilities of a searcher’s needs. The promise of supplementation stands as a corrective to the danger that something relevant has been passed over. Finally, the measure points in the direction of future research for presently unrecognized semantic relationships and toward presently unrecognized connections between terms and concepts that could form new combinations of meaning. The property of relevance conditions both the possible and the actual.

### **Concluding remarks**

For over half a century, precision and recall measures have exercised a powerful influence on the imagination of information retrieval researchers. Studies of precision ratios have focused attention on synonymic relationships and improving keyword searching techniques. Studies of recall ratios purport to measure a system’s ability to return items relevant to a searcher’s interest among all such items in a collection or data structure. The concept of relevance is the criterion by which the success of the measures is judged. Relevance judgements are by their very nature subjective and are dependent upon contextual factors. The determination

of possibly relevant items not retrieved is problematic and can give the impression of circular reasoning. Although subject to difficult questions, understanding the recall denominator as a dependent part that functions analogically clarifies the supplemental nature of the measure's contribution to information retrieval theory. The intuition of relevant items that have yet to be collected serves as a guiding light toward the fulfilment of a searcher's needs.

It is beyond the scope of this article to discuss current trends in information retrieval research or to consider more complex methods of search engine evaluations. The purpose has been to show, by means of an ontological analysis, the structural components that have made these earlier metrics historically appealing. Special attention has been devoted to the problems associated with recall. Viewed inductively, the recall measure has serious difficulties when applied to large, uncontrolled data structures and is subject to the *petitio principii* fallacy. Alternatively, the interaction of the measure's components is intriguing and suggestive of new discoveries. Relationships such as part-whole and dependent-independent are subtly entrenched in our ways of thinking, arising as they do from our experiences as embodied, perceptually oriented creatures. The recognition of structurally dependent parts is closely tied to the successful operation of logical method and reasoning. Despite the many difficulties, what keeps the recall measure enigmatic is the fact that the possibles could become real. They are the objects of an intellectual process.

## Notes

- 1 Considering this point in the context of thesaural relationships, it is clear that possibly relevant items not retrieved could not be represented by terms in the broader term/narrower term hierarchies. Such terms designate membership in a concept class and, as such, would be immediately recognized as relevant by an automated information retrieval system. The possibly relevant items would instead have to be represented by related terms in an associative or analogical relationship to the query topic, terms that are not recognized as relevant by an indexer or an automated indexing mechanism.
- 2 Understanding the role of possibly relevant items in this way is reminiscent of other ontological doctrines in the Western tradition. The Franciscan philosopher Duns Scotus held that items are really distinct from each other if they can be separated, even if it requires a supernatural agent to separate them. Otherwise, they are in fact identical. And, yet, within one real thing, it is possible to distinguish objective components that are different but cannot be separated. These inseparable, distinct components or parts are not mental fictions. The existence of formally distinct components or parts within the same complex entity provides a real ground for applying different concepts to the entity itself. Distinctions that have an objective basis of this kind were called by him "formal distinctions" (*distintio formalitates*).

## References

- Baeza-Yates, Ricardo, and Berthier Ribeiro-Neto. 2011. *Modern Information Retrieval: The Concepts and Technology behind Search*. 2nd ed. New York: Addison Wesley.

- Casati, Roberto, and Achille C. Varzi. 1997. "Spatial Entities." In *Spatial and Temporal Reasoning*, ed. Oliviero Stock, 7–9. Dordrecht: Kluwer Academic Publishers. [http://dx.doi.org/10.1007/978-0-585-28322-7\\_3](http://dx.doi.org/10.1007/978-0-585-28322-7_3).
- Chandrasekaran, B., J. R. Josephson, and R. Benjamins. 1999. "What Are Ontologies, and Why Do We Need Them?" *IEEE Intelligent Systems* 14 (1): 20–26. <http://dx.doi.org/10.1109/5254.747902>.
- Fine, Kit. 1995. "Ontological Dependence." *Proceedings of the Aristotelian Society* 95 (1): 269–90. <http://dx.doi.org/10.1093/aristotelian/95.1.269>.
- Gómez, M., and Chema Abasolo, 2003. "A General Framework for Meta-Search Based on Query-Weighting and Numerical Aggregation Operators." In *Intelligent Systems for Information Processing: From Representations to Applications*, ed. B. Bouchon-Meunier, L. Foulloy, and R. R. Yager, 130–31. Amsterdam: Elsevier. <http://dx.doi.org/10.1016/B978-0-444-51379-3/50010-8>.
- Hurley, Patrick. 2014. *A Concise Introduction to Logic*. 12th ed. Belmont, CA: Wadsworth.
- Jansen, Bernard J. 2009. *Understanding User-Web Interactions via Web Analytics: Synthesis Lectures on Information Concepts, Retrieval, and Services*. Williston, VT: Morgan and Claypool.
- Kowalski, Gerald. 2011. *Information Retrieval Architecture and Algorithms*. New York: Springer. <http://dx.doi.org/10.1007/978-1-4419-7716-8>.
- Lampert, Jay. 1995. *Synthesis and Backward Reference in Husserl's Logical Investigations*. Dordrecht: Kluwer Academic Publishers. <http://dx.doi.org/10.1007/978-94-015-8443-2>.
- Pribbenow, Simone. 1999. "Parts and Wholes and Their Relations." In *Mental Models in Discourse Processing and Reasoning*, ed. Gert Rickheight and Christopher Abel, 368–70. Amsterdam: Elsevier. [http://dx.doi.org/10.1016/S0166-4115\(99\)80059-4](http://dx.doi.org/10.1016/S0166-4115(99)80059-4).
- Raybeck, Douglas, and Douglas Herrmann. 1990. "A Cross-Cultural Examination of Semantic Relations." *Journal of Cross-Cultural Psychology* 21 (4): 452–73. <http://dx.doi.org/10.1177/0022022190214004>.
- Rescher, Nicholas. 1979. "The Ontology of the Possible." In *The Possible and The Actual*, ed. Michael J. Loux, 173–75. London: Cornell.
- Rowley, J., and Richard Hartford. 2008. *Organizing Knowledge: An Introduction to Managing Access to Information*. 4th ed. Farnham, UK: Ashgate.
- Wierzbicka, Anna. 1994. "Semantic Primitives across Languages." In *Semantic and Lexical Universals: Theory and Empirical Findings*, ed. C. Goddard and A. Wierzbicka, 488–90. Philadelphia: John Benjamins. <http://dx.doi.org/10.1075/slcs.25.22wie>.
- Winston, M. E., R. Chaffin, and D. Herrmann. 1987. "A Taxonomy of Part-Whole Relations." *Cognitive Science* 11 (4): 417–44. [http://dx.doi.org/10.1207/s15516709cog1104\\_2](http://dx.doi.org/10.1207/s15516709cog1104_2).