Calculating Salience of Knowledge

Lisa F. Rau
GE Research and Development Center
Schenectady, NY 12301 USA
and
Computer Science Department
University of Exeter, Exeter, UK

Abstract

As information systems continue to grow in size and scope, advances in data management become more and more on the critical path for usability of these systems. This paper reports on the implementation and applicability of an important function - that of calculating the conceptual *salience* of knowledge or data in a knowledge base or database.

Salience is calculated with a method based on Tversky's formulation of salience as composed of two factors: intensity and discriminability. The salience computation has been implemented and tested on a database and is independent of the particular knowledge area.

Introduction

This paper reports on a theory and implementation of the cognitive notion of conceptual *salience*, a concept not typically modelled from a computational perspective. While the concept of *salience of knowledge* has a clear intuitive meaning, this work aims to formalize the notion and provide a computational mechanism for its determination.

Salience of knowledge is, intuitively, the prominence or conspicuousness of knowledge. It is important from a practical perspective because salient knowledge is typically buried along with insignificant knowledge in a large database system. Potentially important facts and relationships are represented in the same way as unimportant information. Discovering what is salient *adds* knowledge of hitherto unknown relationships that can, in turn, be used to reason with and increase the utility of the data represented. Moreover, salient knowledge should be more *accessible* and more *relevant* to a given task, and should be chosen preferentially over more obscure facts.

Salience is a characteristic of knowledge that is important for case-based reasoning [Seifert, 1989], analogical reasoning [Gentner, 1983], knowledge discovery in databases [Piatetsky-Shapiro and Frawley, 1991], understanding metaphor [Makoto *et al.*, 1990] and information retrieval in general. Knowing what characteristics of knowledge are salient allow processes to only deal with those characteristics. Salience, however, is not typically encoded at the time the knowledge is entered, and in fact, salience can only be computed with respect to the other knowledge in the system. The salient portions of a database are in effect only the highlights.

This paper reports on a system that automatically extracts information from text, stores it in a database and discovers salient features. Although the methods are applicable to an arbitrary database or knowledge base, the origins of this database from a real-world source are suggestive of their use in automatically computing the salience of general or specific knowledge.

This work is one portion of a larger project to provide computational methods for automatically deriving what a system knows, for example, in terms of its *breadth of knowledge* [Rau, 1992]. The other "meta-properties" of knowledge are computed based on similar computational foundation as the salience computation.

Overview of Methodology

This paper reports on the theory, implementation and testing of a method for computing what, in an arbitrary knowledge or database, is salient. This section describes both the intuition behind the formalization, and the formalization itself. Following sections describe the database used in demonstrating the implementation, and detail the results.

Tversky [Tversky, 1977] hypothesized that salience is composed of two factors: (1) the intensity of an aspect of a concept (the amount of information), which we denote $\mathcal{I}$, and (2) discriminability, denoted $\mathcal{D}$, corresponding to how well an aspect of a concept distinguishes that concept from related concepts. Iwayama [Makoto *et al.*, 1990] proposed a method, using information theory, to compute and combine these two factors. The method uses a probabilistic model of conceptual categorization. With this representation, the intensity of a concept is equal to its information theoretic redundancy (the inverse of the Shannon [Shannon and Weaver, 1949] entropy); a measure of the amount of uncertainty present in the frequency distribution of values. The discriminability is the ratio of this intensity to the sum of related concepts' intensities. The exact computation is detailed below.

This computational method has been extended to uncover salient combinations of features in a database. This is accomplished by taking all binary combinations of database fields, and computing the salience of the fillers of those fields with respect to related fillers.

Probabilistic Model of Conceptual Categorization

Underlying the calculation of salience is a probabilistic model of conceptual categorization [Smith and Medin, 1988]. Given a database composed of fields $\mathcal{F}$ that contain fillers $\mathcal{F}_i$, we treat the fields as *concepts* or *conceptual categories* and the fillers as *features* or *aspects* of those concepts. Then the frequency of occurrence of each filler in the database approximates the frequency of occurrence of features of a conceptual category. Note that this assumption limits the salience computation to what is salient with respect to the area of expertise in a database.

For example, suppose a database field of **sex-of-person** contained two fillers with the following frequencies of occurrences:

sex-of-person = {(male, .5) (female, .5)}

We denote these probabilities p_i. The category of **sex-of-person** is assumed to be composed of and defined by two features, male and female, each of which occurs with equal probability. On the other hand, a database from a medical office that deals exclusively with pregnancy would have a different defining notion of what sex the patients were. This is important as what is salient to a particular database is necessarily dependent on the particular context and bias of that database.

Amount of Information

Taking the database field to be $\mathcal{F}$ and the individual fillers to be the $\mathcal{F}_i$, the amount of information is the normalized inverse of the well-known Shannon [Shannon and Weaver, 1949] measure of entropy $E(F_i)$:

$$E(F_i) = -\sum_{j=1}^{m} p_{i,j} \log_2 p_{i,j}$$

This entropy measure is adjusted to reflect the total number of distinct fillers (m) by dividing by a normalization factor $\log_2 m$ to obtain relative entropy $e(F_i)$:

$$e_i = \begin{cases} 0 & \text{if } m = 1 \\ \frac{E(F_i)}{\log_2 m} & \text{otherwise} \end{cases}$$

The amount of information varies inversely with the relative entropy, so we define the amount of information $\mathcal{I}$ to be $I = 1 - e(F_i)$. The amount of information is zero when all values are equiprobable, and is one when all values are the same, i.e., when there is only one value for the field. Intuitively, the amount of information measures the variability in frequency of occurrence among different fillers of a field. If all the fillers occur with roughly equal frequencies, than no one filler "stands out" from the rest, hence this component of the salience is low.

Example

To make the calculation concrete, we calculate an example from the domain of experimentation; incidents involving terrorism in Latin America. We calculate the amount of information of **location-of-kidnapping**. Note that we could also calculate the amount of information of **location-of-incident-types** in general (a category that includes other incidents such as **bombing** and **attack**. However the computation of salience takes *slices* of the database, looking at the distribution of fillers with respect to a particular value. We denote the distribution of fillers with respect to a particular value $\mathcal{S}_i$, as opposed to $\mathcal{F}_i$. This is discussed in the next section.

The field of **location** has nine possible fillers, appearing below. To compute the discriminability of **kidnapping** with respect to the location of the incident, we first generate the frequency list of this slice of the database; what proportion of the **kidnappings** occurred with respect to each of the nine **locations**; there were 119 total. This yields:

location-of-kidnapping

	p	-plogp		p	-plogp
Colombia	.50	.50	Venezuala	.01	.07
ElSalvador	.21	.47	Peru	.04	.19
Guatemale	.19	.46	Ecuador	.02	.11
Chile	.02	.11	Brazil	.02	.11
Panama	.01	.07			

The most frequent value is **COLOMBIA**. Under the formula given above for amount of information we can compute: $\mathcal{I}(location - of - kidnapping) = 1 - \frac{2.08}{log_2 9} = 1 - \frac{2.08}{3.17} = 1 - .66 = .34$

Discriminability

In order to compute the measure of discriminability $\mathcal{D}$, there must be a notion of what the concept is to be differentiated with respect to. In database terms, it is possible to look at the variation in the distribution of fillers of a given field with respect to a different field. For example, a database that contained people's occupations and education levels could determine how differentiating a given occupation is with respect to education level, or how good a differentiator or discriminator a certain education level is with respect to occupation. The relation is not symmetric because the measure of discriminability incorporates the amount of information of related concepts; this measure is different for occupation than for education level. We call the distribution of fillers of a field with respect to a different field a *slice* of the database.

The discriminability is calculated by taking the ratio of the amount of information of a slice to the sum of all the amount of information of slices of related concepts. Only related concepts that have the same most frequently occurring filler contribute to this sum.

Continuing with the example, to compute the discriminability, we look at the similar concepts to **kidnapping** that have the same most frequent value of **location**, in this case, **COLOMBIA**. This entails computing the frequency distribution of each slice of incident-type with respect to location. There is only one similar concept that has this same most frequent value of location; **BOMBING** and its amount of information is **.36**. Hence the discriminability is:

$$.34/(.36 + .34) = .34/.70 = .48$$

The ratio of the amount of information of location-of-kidnapping to location-of-incident is the discriminability, in this case, **.48**. This value ranges from nearly zero to one.

This value approaches zero when the denominator is large, which corresponds to when there are many similar concepts with this most frequent value. This value is one when no other similar concepts have this most frequent value, in which case the numerator and denominator are the same.

Salience

Finally, the salience is obtained by multiplying the two terms, $\mathcal{I}$ and $\mathcal{D}$. In the example of **location-of-kidnapping**, the salience is simply:

$$.34 \times .48 = .16$$

This reflects the contribution of two factors; the amount to which the filler and field combination discriminates among similar concepts, and the inherent *amount of information* that that filler has with respect to its field. If it is no more common than any other filler, the amount of information is very low or zero, thus diluting this filler's salience. Conversely, if it is the only filler that field has, the amount of information conveyed by that filler is a maximum.

The limiting cases cannot be determined by more simple counting measures. For example, it can be easily determined (1) when there is only a unique value for any given database slice and (2) when the most frequent value does not occur in any other similar concept. However, the combination of (1) and (2) is extremely unlikely, and for (2) the information theoretic redundancy (amount of information) is equal to the salience, as the discriminability is equal to one (numerator and denominator are equal). In this case, the salience is equal to the amount of information inherent in that data slice. Salience, therefore, combines discriminability with information content. Features that are highest in both will be the most salient. In practice, fillers that occur with very low frequencies (once or twice) tend to give high salience results, however they are typically errors or anomalies. The demonstration performed here excluded fillers that occurred less than twice. Examples are given in the next section.

Implementation

In order to determine what in the database is salient, the system examines all possible slices and computes the relative salience of each slice with respect to various fillers. After inverting a database with n distinct fillers, (an $\mathcal{O}(n)$ operation), it is computationally tractable to compute intersections of pairs of fillers (an $\mathcal{O}(n^2)$ operation). Holding one field constant, the system looks at the distribution of fillers with respect to the constant field, and its related fields. This allows the system to make relative comparisons between, say, the pattern of **instrument** used among different **countries**. In this case, the particular filler of the **country** field is kept constant, and the distribution of values of the instruments within that country are examined.

For every pair of fields, the amount of information, salience and discriminability are calculated and classified according to the following categories:

Discriminability of One: A discriminability of one indicates that the numerator (the amount of information) and the denominator (the amount of information of the related concepts that have the same most likely values) are equal. This means that no other related concept has the same most likely property as the numerator concept. In this case, the salience is equal to the amount of information. If there are many roughly equally frequently occurring fillers, the salience is low.

Amount of Information of One: An amount of information value equal to one indicates that there is only *one value* for this value with respect to this field. In this case, the discriminability and the salience are identical.

Salience of One: If all three values are one, then there is *only one value* for this combination of value with respect to field, and *no other related field has the same most likely property*. Intuitively, that singular value distinguishes among related slices. If the amount of information is one and the discriminability is zero, which implies a zero salience, it is the case that the slice is empty, see below under **Identical**.

Amount of Information of Zero: An amount of information of zero means that all values occur with equal values, that is the frequency of occurrence of all the values is identical. In this case, we define the discriminability to be NIL, and the salience is at a minimum of zero as well. A salience of zero may also indicate that the product of amount of information with the discriminability is very small; so a check for a null discriminability is necessary to distinguish this case.

Ordinary: If none of the above conditions hold, there is at least one other related concept that has the same most likely value. This is the typical case, and here the salience and discriminability are just as defined.

In the cases where more than one value occurs with the same maximum frequency, the amount of information, salience and discriminability are calculated for each of these most likely values.

Demonstration

This section describes the database used to demonstrate the methods just described. The database used contained almost 2,000 database records, each of which has 24 fields of information. The highest accuracy records were manually created to be used to test the accuracy of automated methods of data extraction [Krupka *et al.*, 1991], and it was these manually created records that were used in this demonstration. The fields that contains strings of natural language were made canonical (and conceptual) by running them through the same natural language program that generates the entire templates. These records were created from texts reporting on terrorist activities in Latin America, and we have natural language text processing programs described elsewhere [Jacobs and Rau, 1990; Jacobs and Rau, 1993] capable of generating these records with close-to-human accuracy. Using news stories as a source suggests that this work has the potential to operate on arbitrary and general knowledge, as well as specific databases. Figure 1 shows a sample message and template from this set.

Results of Demonstration

This section describes the major results of the demonstration. 22,320 salience measurements were computed by looking at all slices of the database that contained over two members. From these, the top scoring results are reprinted here. The most salient slices that contained null values are not included in this summary, although the salience of slices that contained null values was computed. The correlations with missing information can be useful, but they convey less information than the other associations.

In what follows, the `Magnitude` is the number of times this combination occurred in the database,

DEV-MUC4-0351

BOGOTA, 18 AUG 89 (EFE) -- [TEXT] SENATOR LUIS CARLOS GALAN, LIBERAL PARTY PRESIDENTIAL HOPEFUL, WAS SHOT THIS EVENING WHEN HE WAS ABOUT TO GIVE A SPEECH AT MAIN SQUARE OF SOACHA, 15 SOUTH OF BOGOTA, IT WAS CONFIRMED BY POLICE AND HEALTH AUTHORI

ACCORDING TO THE FIRST REPORTS, AT LEAST ONE MAN FIRED ON THE SEN TOR FROM AMONG THOSE GATHERED. THE SENATOR IS CURRENTLY AT TH EMERGENCY ROOM OF A HOSPITAL IN BOSA, CLOSE TO SOACHA. TWO OTH PERSONS WERE WOUNDED DURING THE ATTACK.

0. MESSAGE: ID	DEV-MUC3-0351
1. MESSAGE: TEMPLATE	1
2. INCIDENT: DATE	18 AUG 89
3. INCIDENT: LOCATION	COLOMBIA: SOACHA (CITY)
4. INCIDENT: TYPE	ATTACK
5. INCIDENT: STAGE OF EXECUTION	ACCOMPLISHED
6. INCIDENT: INSTRUMENT ID	-
7. INCIDENT: INSTRUMENT TYPE	GUN: "-"
8. PERP: INCIDENT CATEGORY	TERRORIST ACT
9. PERP: INDIVIDUAL ID	"AT LEAST ONE MAN" / "ONE MAN'
10. PERP: ORGANIZATION ID	
11. PERP: ORGANIZATION CONFIDENCE	-
12. PHYS TGT: ID	*
13. PHYS TGT: TYPE	*
14. PHYS TGT: NUMBER	*
15. PHYS TGT: FOREIGN NATION	*
16. PHYS TGT: EFFECT OF INCIDENT	*
17. PHYS TGT: TOTAL NUMBER	*
18. HUM TGT: NAME	"LUIS CARLOS GALAN"
19. HUM TGT: DESCRIPTION	"LIBERAL PARTY PRESIDENTIAL HC "SENATOR": "LUIS CARLOS GALAN "TWO OTHER PERSONS"
20. HUM TGT: TYPE	GOVERNMENT OFFICIAL: "LUIS CA GALAN" CIVILIAN: "TWO OTHER PERSONS"
21. HUM TGT: NUMBER	1: "LUIS CARLOS GALAN" 2: "TWO OTHER PERSONS"
22. HUM TGT: FOREIGN NATION	-
23. HUM TGT: EFFECT OF INCIDENT	INJURY: "LUIS CARLOS GALAN" INJURY: "TWO OTHER PERSONS"
24. HUM TGT: TOTAL NUMBER	

Figure 1: Example Text and Data Extracted

the `Salience` is the actual numerical salience of the result. Recall that the slices compute frequency distributions of database fields with respect to a particular filler, so that that `Filler-2` comes from a different database field than the `Field-1`. The `Field-1` and `Filler-2` define the slice of the database, where `Filler-2` comes from `Field-2`. The `Most Likely Value` is the most frequently occurring filler in this slice; it is one of the fillers of the `Field-1`. For example, the first result indicates that the salience of `human-effect-of-accomplished` is .531, and that the most likely human effect when the event is accomplished is `DEATH`. Recall in the earlier example, we calculated the salience of `location-of-kidnapping` where the most likely value was also `COLOMBIA`. All combinations that appeared over 50 times are shown and with a salience of over .2 are shown.

Analysis

It is always a difficult problem to evaluate automated discovery systems - the human cannot determine what discoveries were not found, and there are no general methods of judging the inherent goodness of any given discovery. However it is safe to say that the relationships categorized as "salient" by this method indeed serve to discriminate among related concepts, and are prominent in terms of relative frequency of co-occurrence when compared to other similar data slices.

Field-1	Filler-2	Field-2	Most-Likely-Value	Magnitude	Salience
HUMAN-EFFECT	ACCOMPLISHED	STATE	DEATH	487	.5310
PHYSICAL-EFFECT	ACCOMPLISHED	STATE	SOME-DAMAGE	212	.5110
LOCATION	TERRORIST-ACT	CATEGORY	COLOMBIA	335	.4710
PERPETRATOR-ORG	PERU	LOCATION	SHINING-PATH	53	.4460
INSTRUMENT-TYPE	ATTACK	TYPE	GUN	123	.4090
INSTRUMENT	BOMBING	TYPE	BOMB	155	.4070
PERPETRATOR-ORG	EL-SALVADOR	LOCATION	FMLN	155	.3910
LOCATION	ACCOMPLISHED	STATE	EL-SALVADOR	424	.3450
INSTRUMENT	BOMBING	TYPE	BOMB	108	.3440
LOCATION	SHINING-PATH	PERPETRATOR-ORG	PERU	53	.3330
HUMAN-TYPE	ACCOMPLISHED	STATE	CIVILIAN	502	.3240
PERPETRATOR-ORG	TERRORIST-ACT	CATEGORY	FMLN	158	.3200
INSTRUMENT-TYPE	TERRORIST-ACT	CATEGORY	BOMB	136	.3140
HUMAN-EFFECT	SOME-DAMAGE	PHYSICAL-EFFECT	INJURY	61	.3060
PHYSICAL-EFFECT	TERRORIST-ACT	CAT	SOME-DAMAGE	92	.2980
HUMAN-EFFECT	BOMBING	TYPE	INJURY	86	.2970
INSTRUMENT	TERRORIST-ACT	CATEGORY	BOMB	96	.2720
CATEGORY	ACCOMPLISHED	STATE	TERRORIST-ACT	790	.2140
INSTRUMENT	ACCOMPLISHED	STATE	BOMB	148	.2020

Figure 2: Sample of Results of Salience Computation

Uses

The salience result can be used in a variety of different ways. Some of the high-salience data reflect logical associations between slots, such as that when a **PHYSical-TARGET** suffers **NO DAMAGE**, any **HUMAN-EFFECT** is likely to be **NO INJURY OR DEATH**. Another example of these logical associations is the relationship between certain perpetrator organizations (**PERP-ORG**, for example **SHINING PATH**) and the location **PERU** where these organizations reside. Detecting such slot inter-dependencies is critical in order to correctly apply any future machine learning methods that assume independence.

Another use of these results is to aid in the determination of which questions to ask to effectively differentiate an event. For example, suppose an analyst is interested in discriminating **TERRORIST ACTs** from **STATE-SPONSORED VIOLENCE**. The most effective slot to know is that which is most salient with respect to the slot one wishes to differentiate upon. This gives the analyst guidance as to which information is most likely to differentiate one from the other. For example, in this case, the **PERPetrator-ORG** differentiates these two types of events very well. The salience results also allow a system analyst to make predictions. For example when a bombing of an **ENERGY** structure is encountered, the above results lend credence to the hypothesis that it was **DESTROYED** in **EL SALVADOR** and that it was a **TERRORIST ACT**. This prediction is justified because these fillers occur more frequently than any other, and discriminate between other types of structures that are **PHYSical-TARGETs**.

Related Work

This computation of salience builds upon the original formulation of Tversky [Tversky, 1977] and the implementation outlined in Iwayama, et. al. [Makoto *et al.*, 1990]. In particular, this paper expands the applicability to an arbitrary database by abandoning the distinction between concepts and features of concepts. We examine all combinations of fillers and fields exhaustively. That is, Iwayama assumes there are categories such as **fruit**, with members such as **apples, lemons**, and that these category members have certain *attributes* or *features* such as **color** and **shape**. Here, we propose one database field (say **incident-type**, which has the fillers such as **kidnapping** and **bombing**) to be a category and examine the salience with respect to another database field (say **location**), putting the **location** field in place of the attribute or role. We also compute the salience of the reverse situation, examining all locations with respect to the incident-types that occur with those locations. While it may seem as if fruit somehow "makes a better category" than color, and color makes a better "role" than fruit, in fact these distinctions are artificial. It is possible to compute the salience of **color-of-apple** just as it is possible to compute the salience of **fruit-of-red**. In the one case, apples are fillers of fruit categories that have color roles. In the other case, red is a filler of color categories that have fruit roles. This blurring of the distinction between categories and roles enables the determination of what is most salient in an arbitrary database. Finally, we have shown the utility of the measure by running it on a real database where the frequency, values are empirically determined.

A great deal of research has addressed the problem of what a system might know or believe [Halpern, 1986; Vardi, 1988]. The work described here contributes to that body of research by adding a new metric that is calculated from what is known, the salience of knowledge. This work is related to recent work in the area of knowledge discovery in databases [Piatetsky-Shapiro and Frawley, 1991] that attempts to learn new knowledge from the structure and content of databases. However, the particular problem of computing salience of knowledge has not been directly addressed in this new research area.

Limitations and Future Directions

The primary area for future work is in the application of the techniques described here to improve the efficiency and accuracy of real programs. Some possibilities are to focus search on salient items, relax an information request for case-based reasoning or information retrieval along salient dimensions, filter out salient discoveries from the output of a machine learning program, and focus reasoning processes on salient characteristics of a problem domain.

One theoretical issue still to be investigated here is the effect of context on the set of "related items" used in this salience computation. As has been shown by Ortony [Ortony *et al.*, 1985], the features of concepts and concepts themselves judged as similar (related) is heavily influenced by context. One artifact of this implementation has to do with individual styles of creating the answer key from which that data was obtained. Each participant created 100 templates, and some had particular ways of indicating certain events that other sites did not. This makes the peculiarities of a given individual's template filling style appear salient. This artifact can be an advantage in that the methods described here can detect such pecularities to improve the consistency of any database where data is entered manually by a variety of individuals.

Conclusions

This paper began with an analysis of the notion of conceptual salience. A specific computation was detailed for automatically determining the conceptual salience of a knowledge or database. The computation combines the amount of information with the discriminability to produce a numerical score. This calculation was validated by computing the salience of combinations of database fillers on a 1,900 record database.

This work is important not only for the methods and computations described here, but for investigating new questions we would like large knowledge based system to be able to answer - questions such as "what do you know that is important?" and "what stands out?". Looking at areas traditionally reserved for the purely cognitive realm, such as meta-questions of knowledge scope and extent, offers a new perspective from which to develop computational answers.

References

[Gentner, 1983] D. Gentner. Structure-mapping: A theoretical framework for analogy. *Cognitive Science*, 7(2), 1983.

[Halpern, 1986] J. Y. Halpern, editor. *Proceedings of the First Conference on Theoretical Aspects of Reasoning About Knowledge*. Morgan Maufmann, Los Altos, CA, March 1986.

[Jacobs and Rau, 1990] P. S. Jacobs and L. F. Rau. The GE NLToolset: A software foundation for intelligent text processing. In *Proceedings of the Thirteenth International Conference on Computational Linguistics*, volume 3, pages 373–377, Helsinki, Finland, 1990.

[Jacobs and Rau, 1993] P. S. Jacobs and L. F. Rau. Innovations in text interpretation. *Artificial Intelligence (Special Issue on Natural Language Processing)*, 48, To Appear 1993.

[Krupka *et al.*, 1991] G. R. Krupka, P. S. Jacobs, L. F. Rau, and L. Iwańska. Description of the GE NLToolset system as used for MUC-3. In *Proceedings of the Third Message Understanding Conference (Muc-3)*, San Mateo, Ca, May 1991. Morgan Kaufmann Publishers.

[Makoto *et al.*, 1990] I. Makoto, T. Takenobu, and T. Hozumi. A method of calculating the measure of salience in understanding metaphors. In *Proceedings of the American Association for Artificial Intelligence*, pages 298–303, Los Altos, CA, August 1990. AAAI, Morgan Kaufmann, Inc.

[Ortony *et al.*, 1985] A. Ortony, R. J. Vondruska, M. A. Foss, and L. E. Jones. Salience, similes, and the asymmetry of similarity. *Journal of Memory and Language*, 24:569–594, 1985.

[Piatetsky-Shapiro and Frawley, 1991] G. Piatetsky-Shapiro and W. J. Frawley. *Knowledge Discovery in Databases*. MIT Press, Cambridge, MA, 1991.

[Rau, 1992] L. F. Rau. Calculating breadth of knowledge. In *Proceedings of the Fourteenth Annual Conference of the Cognitive Science Society*, Hillsdale, NJ, July 1992. Cognitive Science Society, Lawrence Erlbaum Associates.

[Seifert, 1989] C. Seifert. Analogy and case-based retrieval. In *Proceedings of the Darpa Case-Based Reasoning Workshop*, pages 125–129, San Mateo, May 1989. Darpa, Morgan Kaufmann Publishers.

[Shannon and Weaver, 1949] C. E. Shannon and W. Weaver. *The Mathematical Theory of Communication*. University of Illinois Press, Urbana, Il, 1949.

[Smith and Medin, 1988] E. E. Smith and D. L. Medin. *Categories and Concepts*. Harvard University Press, Cambridge, Ma, 1988.

[Tversky, 1977] A. Tversky. Features of similarity. *Psychological Review*, 84(4):327–352, 1977.

[Vardi, 1988] M. Y. Vardi, editor. *Proceedings of the Second Conference on Theoretical Aspects of Reasoning About Knowledge*. Morgan Maufmann, Los Altos, CA, March 1988.