

Universidade de São Paulo (USP)
Universidade Federal de São Carlos (UFSCar)
Universidade Metodista de Piracicaba (Unimep)

Relatório Técnico “Métricas de Avaliação”

<http://gbd.dc.ufscar.br>

**Projeto “Um Ambiente para Análise de Dados da
Doença Anemia Falciforme”**

Pablo Freire Matos (UFSCar)
Leonardo de Oliveira Lombardi (Unimep)
Prof. Dr. Ricardo Rodrigues Ciferri (UFSCar)
Prof. Dr. Thiago Alexandre Salgueiro Pardo (USP/ICMC)
Prof^a. Dr^a. Cristina Dutra de Aguiar Ciferri (USP/ICMC)
Prof^a. Dr^a. Marina Teresa Pires Vieira (Unimep)
pablo_matos@dc.ufscar.br, lolombardi@unimep.br, ricardo@dc.ufscar.br,
{[taspardo](mailto:taspardo@icmc.usp.br), [cdac](mailto:cdac@icmc.usp.br)}@icmc.usp.br, mtvieira@unimep.br



<http://sca.dc.ufscar.br>

**São Carlos
Setembro/2009**

RESUMO

Este relatório técnico visa apresentar as principais métricas utilizadas em sistemas de extração de informação como precisão, revocação e *Medida-F* e a medida de concordância *Kappa* utilizada para avaliar a concordância, por exemplo, de algoritmos de aprendizado de máquina e de humanos. Busca-se com este relatório propiciar aos docentes, discentes, pesquisadores e pessoas interessadas em conhecer essas métricas de avaliação.



LISTA DE FIGURAS

Figura 1 – Medidas dependem do resultado da consulta (oval/azul e roxo).....	7
Figura 2 – Exemplo numérico de informação de Complicação (150) e Não Complicação (350).....	11

LISTA DE TABELAS

Tabela 1 – Matriz de confusão de duas classes (Complicação/Não Complicação).	8
Tabela 2 – Matriz de confusão de duas classes (Sadio/Doente).....	9
Tabela 3 – Matriz de confusão de duas classes (Positivo/Negativo).....	9
Tabela 4 – Exemplo numérico de uma matriz de confusão (Complicação/Não Complicação).	11
Tabela 5 – Exemplo de anotação de dez sentenças.	12
Tabela 6 – Exemplo de matriz de confusão 2 x 2.....	12
Tabela 7 – Matriz de confusão genérica.....	13
Tabela 8 – Exemplo de matriz de confusão 3 x 3.....	13
Tabela 9 – Escala de seis níveis de concordância.....	14

LISTA DE ABREVIATURAS E SIGLAS

IA	Inteligência Artificial
RI	Recuperação de Informação

SUMÁRIO

<u>1</u>	<u>INTRODUÇÃO</u>	6
<u>2</u>	<u>MÉTRICAS DE AVALIAÇÃO</u>	6
2.1	MEDIDAS DE DESEMPENHO	6
2.1.1	PRECISÃO E REVOCAÇÃO NO CONTEXTO DA RECUPERAÇÃO DE INFORMAÇÃO	6
2.1.2	PRECISÃO E REVOCAÇÃO NO CONTEXTO DA INTELIGÊNCIA ARTIFICIAL	7
2.1.3	OUTRAS MEDIDAS DE DESEMPENHO	9
2.2	MEDIDA DE CONCORDÂNCIA	11
<u>3</u>	<u>CONSIDERAÇÕES FINAIS</u>	14
	<u>REFERÊNCIAS</u>	15

1 Introdução

Este relatório técnico tem por objetivo descrever as principais métricas de avaliação utilizadas em sistemas de extração de informação e a medida de concordância entre anotadores. Este conhecimento é necessário para os integrantes do projeto “Um Ambiente para Análise de Dados da Doença Anemia Falciforme” entenderem como um sistema de extração é avaliado e como é calculado a concordância entre os anotadores. Este trabalho está sendo desenvolvido em conjunto com a Universidade de São Paulo (Campus de Ribeirão Preto e São Carlos), Fundação Hemocentro de Ribeirão Preto, Universidade Federal de São Carlos e Universidade Metodista de Piracicaba.

2 Métricas de Avaliação

A seguir são explicadas duas métricas de avaliação que se pode utilizar para avaliar e analisar os resultados gerados a partir da Mineração de Textos: medida de desempenho e medida de concordância. Em 2.1 serão apresentadas as medida de desempenho que são comumente utilizadas para avaliar os sistemas. Em seguida, em 2.2, é apresentada a medida de concordância *Kappa* que serve para identificar o quão fácil uma determinada tarefa é realizada pelos seres humanos.

2.1 Medidas de Desempenho

Precisão e revocação são medidas amplamente utilizadas para avaliar a qualidade dos resultados em diversas áreas do conhecimento. Precisão é uma medida de fidelidade, enquanto a revocação (conhecida também como cobertura ou sensibilidade) é uma medida de completude.

As medidas de precisão e revocação são medidas padrão da Recuperação de Informação (RI), Cleverdon (1966 apud SILVA, 2006). As mesmas são utilizadas para contribuir com a avaliação de sistemas de RI que tem o objetivo de recuperar documentos relevantes a partir da consulta de um usuário, porém diversas outras áreas, como Extração de Informação e Inteligência Artificial (IA) incluindo Aprendizado de Máquina e Processamento de Língua Natural, utilizam dessas medidas para avaliação.

Em seguida são explicadas a precisão e a revocação sob o ponto de vista da RI e IA além de outras medidas e os conceitos de falso positivo e falso negativo.

2.1.1 Precisão e Revocação no contexto da Recuperação de Informação

A Figura 1 representa o conjunto de informações de uma coleção de documentos sobre complicações da Anemia Falciforme, discriminadas em informações relevantes (esquerda/verde e azul) e informações irrelevantes (direita/vermelho e roxo). O formato oval (azul/roxo) significa o resultado da consulta sobre informações de complicação. A consulta ideal e desejável é quando for

recuperada toda a parte esquerda da figura (verde e azul), tendo assim 100% de precisão e revocação (equações das medidas explicadas a seguir).

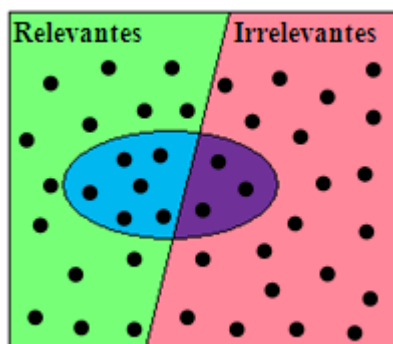


Figura 1 – Medidas dependem do resultado da consulta (oval/azul e roxo).

Na área de RI, a precisão é o número de elementos relevantes recuperados dividido pelo número total de elementos recuperados (Equação (1)) e a revocação é definida como o número de elementos relevantes recuperados dividido pelo número total de elementos relevantes existentes (que deveriam ter sido recuperados) (Equação (2)).

$$\text{Precisão} = \frac{\text{Número de elementos relevantes recuperados}}{\text{Número total de elementos recuperados}} \quad (1)$$

$$\text{Revocação} = \frac{\text{Número de elementos relevantes recuperados}}{\text{Número total de elementos relevantes}} \quad (2)$$

Na RI uma pontuação perfeita de precisão de 1,0 significa que cada resultado obtido por uma pesquisa foi relevante (mas não diz nada sobre se todos os elementos relevantes foram recuperados), enquanto uma pontuação perfeita de revocação de 1,0 significa que todos os elementos relevantes foram recuperados pela pesquisa (mas nada diz sobre quantos elementos irrelevantes também foram recuperados).

Muitas vezes existe uma relação inversa entre precisão e revocação, onde é possível aumentar uma ao custo de reduzir a outra. Por exemplo, um sistema de RI pode aumentar a revocação recuperando mais elementos, ao custo de um número crescente de elementos irrelevantes recuperados (diminuindo a precisão).

2.1.2 Precisão e Revocação no contexto da Inteligência Artificial

A matriz de confusão oferece uma medida efetiva do modelo de classificação, ao mostrar o número de classificações corretas e as classificações previstas para cada classe em um determinado conjunto de exemplos. Em Monard e Baranauskas (2003) encontra-se informações de como preencher uma matriz de n classes.

A Tabela 1 mostra a matriz de confusão para duas classes (Complicação/Não Complicação) da Anemia Falciforme. Nesta tabela, P representa o valor positivo que significa que a palavra-chave

extraída é complicação; N representa o valor negativo que significa que não é complicação; p (Extração de Complicação) e n (Extração de Não Complicação) são, respectivamente, os valores positivo e negativo da extração automática.

Verdadeiro Positivo (VP) significa que uma quantidade X de complicações relacionadas à Anemia Falciforme extraídas do documento é 100% complicação e foi extraída corretamente. Já Verdadeiro Negativo (VN) é o oposto, 100% não é complicação e não foi extraída. Falso Positivo (FP) não é complicação, mas foi erroneamente extraída do documento e Falso Negativo (FN) é complicação, mas não foi precisamente extraída.

Tabela 1 – Matriz de confusão de duas classes (Complicação/Não Complicação).

Condição Atual (Avaliação Especialista)	<i>Complicação (P)</i>	<i>Não Complicação (N)</i>
Resultado da Extração Automática		
<i>Extração de Complicação (p)</i>	VP (Verdadeiro Positivo)	FP (<i>não complicação, mas é extraída</i>)
<i>Extração de não Complicação (n)</i>	FN (<i>complicação, mas não é extraída</i>)	VN (Verdadeiro Negativo)

Incontestavelmente, o resultado que é mais interessante é VP que representa a quantidade de complicações extraídas. Com base nesse resultado é possível saber qual é a precisão e a revocação. A primeira medida calcula a percentagem de acerto a partir das complicações e não complicações que foram extraídas, Equação (3). A segunda calcula a percentagem das complicações que foram extraídas em relação ao total das complicações, Equação (4).

- **Precisão:** Taxa com que todos os exemplos classificados como positivos são realmente positivos. Nenhum exemplo negativo é incluído.

$$Precisão = \frac{VP}{VP + FP} \quad (3)$$

- **Revocação:** Taxa com que classifica como positivos todos os exemplos que são positivos. Nenhum exemplo positivo é deixado de fora. Apresenta uma indicação do quanto do total de informação relevante foi recuperada.

$$Revocação = \frac{VP}{VP + FN} \quad (4)$$

Outro exemplo que pode ser considerado é o diagnóstico de um indivíduo. A classificação é definida como sadio ou doente. Neste caso os falsos positivos são aqueles que estão doentes, mas que foram diagnosticados como sadios e os falsos negativos são os sadios e que foram diagnosticados como doentes. Os verdadeiros positivos são os que estão sadios e são realmente

diagnosticados como sadios e os verdadeiros negativos são os que estão doentes e são diagnosticados como doentes (Tabela 2).

Tabela 2 – Matriz de confusão de duas classes (Sadio/Doente).

Condição Atual		
\	<i>Sadio</i>	<i>Doente</i>
Teste		
<i>Diagnóstico Sadio</i>	Verdadeiro Positivo	FP (<i>doente, mas sadio</i>)
<i>Diagnóstico Doente</i>	FN (<i>sadio, mas doente</i>)	Verdadeiro Negativo

A partir dos exemplos na matriz de confusão ilustrados na Tabela 1 e Tabela 2 é possível extrair outras medidas que possam avaliar o desempenho das complicações extraídas, ou seja, avaliando o quão verdadeira são as informações extraídas. Em seguida são discutidas algumas dessas medidas.

2.1.3 Outras Medidas de Desempenho

Tomando como base as informações contidas na Tabela 3, definiremos as seguintes medidas.

Tabela 3 – Matriz de confusão de duas classes (Positivo/Negativo).

Condição Atual		
\	<i>P</i>	<i>N</i>
Teste		
<i>p</i>	VP	FP
<i>n</i>	FN	VN

Nas Equações (5), (6) e (7) a seguir considera-se $P = \text{Precisão}$ e $R = \text{Revocação}$.

- **Medida-F (F-Measure):** Média harmônica ponderada da precisão e revocação, Equação (5). F_β mede a eficácia da recuperação em relação ao valor atribuído a Beta (β). Pesos comumente utilizados para β são: F_2 (revocação é o dobro da precisão) e $F_{0,5}$ (precisão é o dobro de revocação). A precisão tem peso maior para valores $\beta < 1$, enquanto que $\beta > 1$ favorece a revocação.

$$\text{Medida } F_\beta = \frac{(1 + \beta) \times (P \times R)}{(\beta \times P + R)}, \text{ onde } \beta = \frac{1 - \alpha}{\alpha} \quad (5)$$

A *Medida-F* foi derivada por van Rijsbergen (1979) baseada na medida de eficiência (*effectiveness*), Equação (6), do mesmo autor.

$$E = 1 - \left(\frac{1}{\left(\frac{\alpha}{P} + \frac{(1-\alpha)}{R} \right)} \right) \quad (6)$$

A relação entre a *Medida- F_β* e a medida de eficiência é: $F_\beta = 1 - E$. Quando a precisão e a revocação têm o mesmo peso ($\beta = 1$) a medida é *Medida- F_1* , também conhecida como *Medida- F* tradicional ou *F-Score* balanceada, Equação (7).

$$\text{Medida } F = \frac{2 \times P \times R}{P + R} \quad (7)$$

- **Acurácia:** Mais frequentemente utilizada para avaliação de problemas de classificação de aprendizado de máquina. Há uma boa razão para que esta medida não seja adequada para problemas de RI: em quase todas as circunstâncias, os dados são extremamente desproporcionais e normalmente mais de 99,9% das informações são irrelevantes. Portanto, um sistema cujo objetivo é maximizar a acurácia pode aparentemente ter uma boa avaliação, pois considera todas as informações irrelevantes (MANNING; RAGHAVAN; SCHÜTZE, 2008).

$$\text{Acurácia} = \frac{VP + VN}{VP + VN + FP + FN} \quad (8)$$

- **Especificidade:** Taxa com que uma instância verdadeiramente negativa é classificada como negativa.

$$\text{Especificidade} = \frac{VN}{VN + FP} \quad (9)$$

Para exemplificar o uso dessas medidas, em seguida é mostrado um exemplo de uma matriz de confusão de termos relacionados a complicações da doença Anemia Falciforme com as classes Complicação e Não Complicação (Tabela 4).

Considera-se as seguintes informações fictícias para o preenchimento da Figura 2 e Tabela 4, mostrados a seguir. Suponha que foi realizada uma avaliação manual de uma coleção de documentos relacionados a complicações da Anemia Falciforme por um especialista, que encontrou nessa coleção 150 complicações. Em uma avaliação automática, realizada por um sistema de extração automática, encontrou 160 palavras-chave, sendo 120 complicações e 40 não complicações.

Assim a quantidade de verdadeiro positivo é 120 (é complicação e foi corretamente extraída), de falso positivo é 40 (não é complicação, mas foi extraída erradamente), de falso negativo é $150 - 120 = 30$ (é complicação, mas não foi extraída) e verdadeiro negativo é todo o resto das palavras contidas na coleção de documentos avaliada que não são complicações e não foram extraídas (vamos supor que seja ínfimos 310).

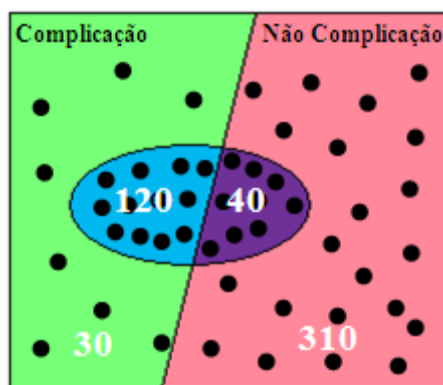


Figura 2 – Exemplo numérico de informação de Complicação (150) e Não Complicação (350).

Tabela 4 – Exemplo numérico de uma matriz de confusão (Complicação/Não Complicação).

Resultado da Extração Automática \ Condição Atual (Avaliação Especialista)	Complicação	Não Complicação
	Extração de complicação	120
Extração de não complicação	30	310

Os valores para as medidas são exemplificados logo em seguida.

$$\text{Precisão} = \frac{VP}{VP + FP} = \frac{120}{120 + 40} = 0,75$$

$$\text{Sensibilidade} = \text{Revocação} = \frac{VP}{VP + FN} = \frac{120}{120 + 30} = 0,8$$

$$\text{Medida } F_1 = \frac{2 \times \text{Precisão} \times \text{Revocação}}{\text{Precisão} + \text{Revocação}} = \frac{2 \times 0,75 \times 0,8}{0,75 + 0,8} \cong 0,77$$

$$\text{Acurácia} = \frac{VP + VN}{VP + VN + FP + FN} = \frac{120 + 310}{120 + 310 + 40 + 30} = 0,86$$

$$\text{Especificidade} = \frac{VN}{VN + FP} = \frac{310}{310 + 40} \cong 0,88$$

2.2 Medida de Concordância

O coeficiente *Kappa* é uma medida de concordância estatística formulada por Jacob Cohen (1960). Adequada para tarefas de classificação realizada por vários anotadores/juízes. Os anotadores – podem ser humanos ou não (algoritmos de classificação é um exemplo de anotador que não seja humano (FIOL; HAUG, 2009)) – têm a função de definir a que classe os exemplos

pertencem, que por sua vez são utilizados como material de treinamento para aprendizado supervisionado (CARLETTA, 1996).

O resultado obtido com a medida *Kappa* define um limiar de concordância entre anotadores seja em tarefas de classificação ou de extração de informação. Esse resultado depende do número de anotadores, número de classes e número de exemplos a serem classificados. O limiar de concordância pode ajudar de três formas:

- Avaliar o desempenho do sistema comparando o limiar em percentual com os percentuais obtidos com as medidas como precisão/revocação;
- Contribuir em avaliar se o conjunto de treinamento é um material válido, sendo, às vezes, necessário mais treinamento para aumentar a concordância;
- Descartar exemplos controversos. Exemplos que não sejam fáceis dos anotadores concordarem podem contribuir com o aumento do resultado das medidas de desempenho utilizadas.

Na Equação (10) é mostrada o coeficiente da medida *Kappa* (k) ou simplesmente índice *Kappa*, onde $P(A)$ é a proporção de vezes que os anotadores concordam, Equação (11), e $P(E)$ é a proporção de vezes que é esperado dos anotadores concordarem ao acaso ou aleatoriamente.

$$k = \frac{P(A) - P(E)}{1 - P(E)} \quad (10)$$

$$P(A) = \frac{\text{número de vezes que o anotadores concordaram}}{\text{número de itens anotados}} \quad (11)$$

Suponha que dois anotadores são incumbidos de classificar 10 sentenças em duas classes (complicação e sintoma). O resultado da classificação é mostrado na Tabela 5 e resumido na matriz de confusão na Tabela 6.

Tabela 5 – Exemplo de anotação de dez sentenças.

Sentença	1	2	3	4	5	6	7	8	9	10
Anotador 1	C	S	C	C	C	S	S	C	S	C
Anotador 2	C	S	C	C	S	S	C	C	S	C
Legenda: C = Complicação, S = Sintoma										

Tabela 6 – Exemplo de matriz de confusão 2 x 2.

		Anotador 1		
		Complicação	Sintoma	Total
Anotador 2	Complicação	5 (3,6)	1	6
	Sintoma	1	3 (1,6)	4
	Total	6	4	10

A proporção observada que os anotadores concordaram é: $P(A) = \frac{5+3}{10} \cong 0,8$. Os valores para a distribuição marginal são: $P(\text{Complicação}) = \frac{6 \times 6}{10} = 3,6$ e $P(\text{Sintoma}) = \frac{4 \times 4}{10} \cong 1,6$. A probabilidade dos dois anotadores concordarem por acaso é: $P(E) = \frac{P(\text{Complicação}) + P(\text{Sintoma})}{10} = \frac{3,6+1,6}{10} = 0,52$. A medida *Kappa*, calculada pela Equação (10), é: $k = \frac{0,8-0,52}{1-0,52} \cong 0,58$, ou seja, na tarefa de classificar 10 sentenças em duas classes os humanos, representado por dois anotadores, concordam em 58% das vezes.

O exemplo calculado no parágrafo anterior é para duas classes. Para mais de duas classes considera a matriz de confusão genérica na Tabela 7, onde a matriz de confusão ideal é quando somente a diagonal principal for preenchida e os outros elementos tenham valor zero, ou seja, não obteve nenhuma classificação errada. $P(A)$ é calculado pela exatidão total $\frac{\sum_{i=1}^c X_{ii}}{n}$ e $P(E)$ é calculado pela exatidão total ao acaso, $\frac{\sum_{i=1}^c X_{i+} \cdot X_{+i}}{n^2}$.

Tabela 7 – Matriz de confusão genérica.

		Anotador 1				Total
		1	2	...	c	
Anotador 2	1	x_{11}	x_{12}	...	x_{1c}	x_{1+}
	2	x_{21}	x_{22}	...	x_{2c}	x_{2+}
	⋮	⋮	⋮	⋮	⋮	⋮
	c	x_{c1}	x_{c2}	...	x_{cc}	x_{c+}
	Total	x_{+1}	x_{+2}	...	x_{+c}	n

Considere o exemplo a seguir mostrado na Tabela 8 para três classes quaisquer classificadas por dois anotadores.

Tabela 8 – Exemplo de matriz de confusão 3 x 3.

		Anotador 1			Total
		1	2	3	
Anotador 2	1	25	2	5	32
	2	3	26	5	34
	3	1	3	30	34
	Total	29	31	40	100

O valor de $P(A)$ é $\frac{25+26+30}{100} = 0,81$ e de $P(E)$ é $\frac{32 \times 29 + 34 \times 31 + 34 \times 40}{100^2} \cong 0,33$. Usando a Equação (10) o índice *Kappa* é $k = \frac{0,81-0,33}{1-0,33} \cong 0,72$, ou seja, nesta tarefa os humanos concordam em 72% das vezes.

Segundo Manning, Raghavan e Schütze (2008), há outras duas maneiras de calcular a medida *Kappa* que depende se a combinação da distribuição marginal considera todos os

anotadores ou cada anotador separadamente. O valor da medida varia de $-1..1$, onde $k = 1$ significa concordância completa, $k = 0$ concordância aleatória e $k = -1$ discordância máxima.

Em geral a medida *Kappa* acima de 0,8 é considerada como um bom acordo, entre 0,67 e 0,8 é tido como justo acordo e abaixo de 0,67 é considerado como duvidosa para avaliação, embora a interpretação dos resultados da medida dependa dos fins para os quais os dados serão utilizados (MANNING; RAGHAVAN; SCHÜTZE, 2008). Landis e Koch (1977) propuseram uma escala de seis níveis para interpretar os valores de *Kappa* como mostrado na Tabela 9.

Tabela 9 – Escala de seis níveis de concordância.

Concordância	Escala
Ruim	Abaixo de 0
Leve	0,00 a 0,20
Sofrível	0,21 a 0,40
Moderada	0,41 a 0,60
Boa	0,61 a 0,80
Quase perfeita	Acima de 0,81

3 Considerações Finais

Neste relatório foram discutidas métricas que podem ser utilizadas para avaliar o resultado de algoritmos de classificação e técnicas de extração de informação. Segundo Krauthammer e Nenadic (2004), normalmente o desempenho de sistemas de reconhecimento automático de termo é avaliado em termos de precisão e revocação e o desempenho global é medido pela *Medida-F*.

Referências

- CARLETTA, J. Assessing agreement on classification tasks: the kappa statistic. **Computational Linguistics**, v. 22, n. 2, p. 249-254, 1996. Disponível em: <<http://www.aclweb.org/anthology-new/J/J96/J96-2004.pdf>>. Acesso em: 24 out. 2008.
- COHEN, J. A coefficient of agreement for nominal scales. **Educational and Psychological Measurement**, v. 20, n. 1, p. 37-46, 1960.
- FIOL, G. D.; HAUG, P. J. Classification models for the prediction of clinicians' information needs. **Journal of Biomedical Informatics**, v. 42, n. 1, p. 82-89, 2009. Disponível em: <<http://dx.doi.org/10.1016/j.jbi.2008.07.001>>. Acesso em: 13 mar. 2009.
- KRAUTHAMMER, M.; NENADIC, G. Term identification in the biomedical literature. **Journal of Biomedical Informatics**, v. 37, n. 6, p. 512-526, 2004. Disponível em: <<http://dx.doi.org/10.1016/j.jbi.2004.08.004>>. Acesso em: 25 fev. 2009.
- LANDIS, J. R.; KOCH, G. G. The measurement of observer agreement for categorical data. **Biometrics**, v. 33, n. 1, p. 159-174, 1977.
- MANNING, C. D.; RAGHAVAN, P.; SCHÜTZE, H. **Introduction to information retrieval**. Cambridge: Cambridge University Press, 2008. 482 p. Disponível em: <<http://www-csli.stanford.edu/~hinrich/information-retrieval-book.html>>. Acesso em: 28 nov. 2008.
- MONARD, M. C.; BARANAUSKAS, J. A. Conceitos sobre aprendizado de máquina. In: REZENDE, S. O. (Ed.). **Sistemas inteligentes: fundamentos e aplicações**. São Carlos: Manole, 2003. p. 89-114. cap. 4.
- SILVA, P. P. **ExtraWeb: um sumarizador de documentos Web baseado em etiquetas HTML e ontologia**. 158 f. Dissertação (Mestrado em Ciência de Computação) – Departamento de Ciência da Computação, Universidade Federal de São Carlos, São Carlos, 2006. Disponível em: <http://www.btd.ufscar.br/tde_busca/arquivo.php?codArquivo=1170>. Acesso em: 10 nov. 2008.
- VAN RIJSBERGEN, C. J. **Information retrieval**. 2nd ed. Butterworth-Heinemann, 1979. 224 p. Disponível em: <<http://www.dcs.gla.ac.uk/Keith/Preface.html>>. Acesso em: 10 nov. 2008.