

# RECUPERAÇÃO DE INFORMAÇÃO USANDO COMPUTAÇÃO NEBULOSA A PARTIR DE DOCUMENTOS COM ESTRUTURAS HETEROGÊNEAS

**Gustavo Alberto Giménez Lugo** <sup>⊗</sup>  
Escola Politécnica  
Universidade de São Paulo  
São Paulo - SP  
gustavo@pcs.usp.br

**Marco Túlio Carvalho de Andrade**  
Escola Politécnica  
Universidade de São Paulo  
São Paulo - SP  
mtulio@pcs.usp.br

**Jaime Simão Sichman**  
Escola Politécnica  
Universidade de São Paulo  
São Paulo - SP  
jaime@pcs.usp.br

## RESUMO

*Este trabalho apresenta um experimento que usa técnicas nebulosas para implementar um sistema protótipo para recuperação de informação, reduzindo de  $n$  para 3 o número de seções consideradas em um documento. O objetivo foi verificar o comportamento e o desempenho de tal enfoque quando aplicado a uma coleção de documentos com estruturas heterogêneas, a priori desconhecidas, articulando o texto. Os resultados foram promissores apesar da simplicidade do enfoque, criando um grande interesse na exploração de modelos mais sofisticados de recuperação de informação usando a teoria nebulosa para a construção de ontologias particulares.*

## ABSTRACT

*This work presents an experiment using fuzzy techniques to implement a prototype system for information retrieval, reducing from  $n$  to 3 the number of sections considered in a document corpora. The goal was to verify the behavior and performance of such an approach when applied to a collection of documents with heterogenous, a priori unknown, structures articulating the text. The results were promising, in spite of the rather primitive approach, instilling a deep interest in exploring a more sophisticated information retrieval model, better exploiting fuzzy theory as a way to build particular ontologies.*

## 1 INTRODUÇÃO

A problemática de explosão da informação, hoje tão visada pela pesquisa, já era apontada por Vannevar Bush na década de 40 (Bush, 1945 apud Saracevic, 1995) e é em volta deste ponto que gira o próprio núcleo da recuperação de informação.

A conceituação de Recuperação de Informação, definida inicialmente por Calvin Mooers (Mooers, 1951 apud Saracevic, 1995), vem dada, em geral (Frakes, 1992; Russell e Norvig, 1995; Turtle e Croft, 1997), de uma forma eminentemente funcional e não descritiva:

*Recuperação de informação é o nome do processo ou método onde um possível usuário de informação pode converter a sua necessidade de informação numa lista real de citações de documentos armazenados que contenham informações úteis a ele . . . recuperação de informação abarca os aspectos intelectuais da descrição da informação e a sua especificação para busca, assim como também quaisquer sistemas, técnicas ou máquinas que sejam empregadas para efetuar a operação (Mooers, 1951 apud Saracevic, 1995).*

De qualquer modo, uma vez de posse dos dados da consulta são selecionados os documentos que se apresentam como mais relevantes comparando a consulta com representações dos documentos previamente armazenadas, opcionalmente revisando a representação da consulta para tratar consultas posteriores. O processo de recuperação envolve assim a aquisição da consulta, seleção dos documentos e revisão da consulta. Uma coleção de documentos pode incluir documentos textuais e não textuais, tais como fotografias, som, vídeo, módulos de software, etc.

A explosão de informação é um complexo problema social, cognitivo, cultural e de comunicação e não simplesmente um problema técnico. Isto traz a tona a

visão de Zadeh (Zadeh, 1965; Zadeh, 1973) quanto às vantagens de usar um tratamento nebuloso para lidar com sistemas que interagem com seres humanos. Muito mais ainda quando a própria essência da computação nebulosa está centrada no universo do discurso lingüístico humano, enquanto aquisição e representação do conhecimento. Autores como (Saracevic, 1995) chegam a expressar o sucesso dos sistemas de recuperação de informação em função do problema:

*O quão satisfatório foi e é a Recuperação de Informação pode ser medido pela aplicação na resolução do problema de explosão da informação nas áreas em que foi aplicado?*

Isto leva a outras questões correlatas, como:

*Quão satisfatoriamente a Recuperação de Informação fornece suporte às pessoas nas situações em que elas se defrontam com problemas de pesquisa, achado, uso e interação com informação advinda da massa de informações existentes e a miríade de escolhas possíveis?*

Recuperação de informação é aplicada, dentre outras formas, na busca e navegação através da Internet e constitui o coração da pesquisa e o desenvolvimento das bibliotecas digitais.

Qualquer sistema de recuperação de informação inclui três componentes básicos: identificação e representação do conteúdo do documento, aquisição e representação da necessidade de informação e a especificação da função de comparação que seleciona os documentos relevantes baseado nas representações. A incerteza aparece em cada um dos componentes citados (Croft e Turtle, 1992; Turtle e Croft, 1997).

---

<sup>⊗</sup> Bolsista CNPq