

Ovídio José Francisco

**Aplicação de técnicas de Mineração de Textos
para Organização e Extração de Históricos de
Decisões de Documentos de Reuniões**

Sorocaba, SP

28 de agosto de 2017

Sumário

1	INTRODUÇÃO	3
---	------------------	---

1 Introdução

Reuniões são tarefas presentes em atividades corporativas, ambientes de gestão e organizações de um modo geral. Seu conteúdo é frequentemente registrado em texto na forma de atas para fins de documentação e consulta posterior a qual devida a manual desses arquivos torna-se uma tarefa custosa, especialmente considerando o seu crescimento em uma instituição (??????).

As atas de reunião possuem características particulares, frequentemente têm a característica de apresentar um texto com poucas quebras de parágrafo e sem marcações de estrutura, como capítulos, seções ou quaisquer indicações sobre o tema do texto. O estilo contribui para a escrita de textos sucintos com poucos detalhes, pois o ambiente dá preferência a textos curtos.

Os documentos apresentam um texto com poucas quebras de parágrafo e sem marcações de estrutura, como capítulos, seções ou quaisquer indicações sobre o assunto do texto. É comum a presença de cabeçalhos, rodapés e numeração de páginas e linhas o que pode prejudicar tanto similaridade entre sentenças como a apresentação dos segmentos ao usuário.

As atas são documentos textuais que em geral descrevem dados não estruturados. Assim, um sistema que responde a consultas do usuário ao conteúdo das atas, retornando trechos de textos relevantes à sua intenção, é um desafio que envolve a compreensão de seu conteúdo (??).

Devido a fatores como a não estruturação e volume dos textos, a localização de assunto em uma ata é uma tarefa custosa. Usualmente, o que se faz são buscas manuais guiadas pela memória ou com uso de ferramentas computacionais baseadas em localização de palavras-chave. Normalmente esse tipo de busca exige a inserção de termos exatos e apresentam ao usuário um documento com as palavras buscadas em destaque, mantendo o texto longo, o que dificulta por exemplo o ranqueamento por relevância.

Uma vez que a ata registra a sucessão de assuntos discutidos na reunião, há interesse em um sistema que aponte trechos de uma ata que tratam de um assunto específico. Tal sistema tem duas principais tarefas: 1) Descobrir quando há uma mudança de assunto. 2) Descobrir quais são esses assuntos. Este trabalho tem como foco principal, a detecção de mudança de assuntos, que pode ser atendida pela segmentação automática de textos (??????). E a extração de automática de tópicos.

A tarefa de segmentação textual consiste em dividir um texto em partes que contenham um significado relativamente independente. Em outras palavras, é identificar

as posições nas quais há uma mudança significativa de assuntos.

A segmentação de textos é útil em aplicações que trabalham com textos sem indicações de quebras de assunto, ou seja, não apresentam seções ou capítulos, como transcrições automáticas de áudio, vídeos e grandes documentos que contêm vários assuntos como atas de reunião e notícias.

Pode ser usada para melhorar o acesso a informação solicitada por meio de uma consulta, onde é possível oferecer porções menores de texto mais relevantes ao invés de exibir um documento grande que pode conter informações menos pertinentes. A navegação pelo documento pode ser aprimorada, em especial na utilização por usuários com deficiência visual, os quais utilizam sintetizadores de texto como ferramenta de acessibilidade (??). Além disso, encontrar pontos onde o texto muda de assunto, pode ser útil como etapa de pré-processamento em aplicações voltadas ao entendimento do texto, principalmente em textos longos.