

Avaliação de Técnicas de Recuperação de Informação para Organização e Extração de Conhecimento de Documentos de Reunião

Ovídio José Francisco

Orientadora: Prof.^a Dr. Katti Faceli

Coorientador: Prof. Dr. Rafael Geraldeli Rossi



August 1, 2018

- 1 Contextualização
- 2 Objetivos
- 3 Proposta
- 4 Análise dos Resultados
- 5 Conclusão

As atas registram assuntos discutidos em reuniões e podem ser utilizadas como base de dados.

- Utilizadas como referência e apoio a decisões;
- Um assunto pode ser discutido diversas vezes em reuniões diferentes;
- É desejável recuperar um histórico desses assuntos ao longo do tempo;
- Necessidade de ferramentas automáticas.

Recuperação de Informação em documentos textuais:

- Informações contidas em grandes quantidades de texto;
- Inerentemente não estruturados;
- Documentos com múltiplos assuntos;
- Assuntos dispersos pela base de documentos.

Nesse cenário, o desafio é encontrar trechos de texto que tratem somente do assunto pesquisado.

Essa tarefa consiste em 2 passos principais:

- Encontrar pontos onde há transição de assuntos;
- Identificar o teor desses assuntos;

Os algoritmos de segmentação textual são utilizados para dividir um texto em segmentos contendo um assunto completo e relativamente independente.

- Úteis em aplicações com textos sem indicações de quebras de assunto, como transcrições de áudio, e diálogos em chats.
- Não dão indicações sobre o conteúdo dos segmentos.

Os modelos de Extração de Tópicos podem estimar o assunto de cada documento de uma coleção.

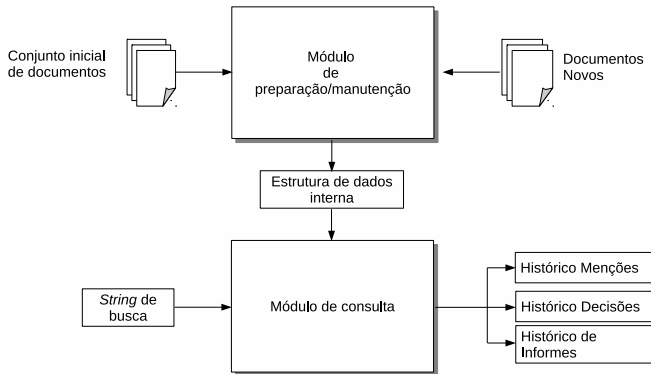
- Agrupam documentos por tópico.
- Identificam palavras para descrever o tópico do documento.

Propor uma solução para identificar, organizar e consultar assuntos registrados em atas de reunião.

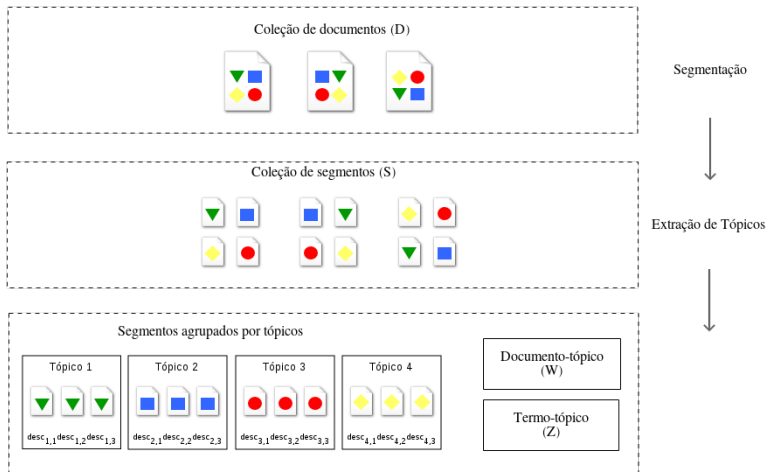
Utilizar técnicas de Segmentação textual em conjunto com modelos de Extração de Tópicos para:

- Gerar uma estrutura mais organizada que a coleção original.
- Utilizar a estrutura latente dos segmentos para Recuperação de Informação.

Visão Geral do Sistema.



Estrutura de dados interna e seu processo de geração.



Estrutura de dados interna e seu processo de geração.



UNIVERSIDADE FEDERAL DE SÃO CARLOS – Campus Sorocab

BACHARELADO EM CIÊNCIA DA COMPUTAÇÃO

Rodovia João Leme dos Santos, Km 110 (SP-264)

Bairro do Itinga – Sorocaba/SP – CEP 13052-780

Telefone: (151) 3202-2022 / www.ufscar.br



UNIVERSIDADE FEDERAL DE SÃO CARLOS – Campus Sorocaba

BACHARELADO EM CIÊNCIA DA COMPUTAÇÃO

Rodovia João Leme dos Santos, Km 110 (SP-264)

Bairro do Itinga – Sorocaba/SP – CEP 13052-780

Telefone: (151) 3202-2022 / www.ufscar.br

Ata da 17ª Reunião Extraordinária do Conselho do Curso de Bacharelado em Ciência da Computação, UFSCar – Sorocaba

Local, Dia e Hora: No laboratório de Pedagogia, situada nas dependências da Universidade Federal de São Carlos – Campus Sorocaba, na Rodovia João Leme dos Santos, quilômetro 110, Bairro do Itinga, na cidade de Sorocaba, Estado de São Paulo e realizada aos vinte e cinco dias do mês de agosto de dois mil e dez, às 14h00.

Convocação e Presença: A convocação foi realizada sete dias antes da data da sua realização, estando presentes os membros do Conselho do Curso de Bacharelado em Ciência da Computação – Sorocaba, sendo eles os professores

representantes discentes Sr.

(2010). **Comunicação da**

Presidência: A prof. informou que na última reunião do ConCam foram indicados os nomes do prof. titular e suplente respectivamente, como representantes do campus Sorocaba no Conselho (Conselho Universitário). A prof. esclareceu também, que o campus já possui um assento ocupado pelo prof. e que agora leremos dois assentos no respectivo conselho. A prof. também informou que na mesma reunião do ConCam foram indicados os nomes do prof. titular e suplente, para o assento do campus Sorocaba no CoAd (Conselho de Administração da UFSCar), destacando que agora o campus Sorocaba não está mais no plano de implantação do campus, e irá concorrer com os outros campi da universidade para distribuição de vagas. A prof. também informou que no próximo dia trinta de agosto haverá uma reunião com o ProGrad em Sorocaba para retorno sobre o Projeto Político Pedagógico do curso no qual a coordenação do curso irá participar.

Comunicação dos membros: O prof. comunicou que houveram diversos problemas com o lançamento do edital para o concurso público da vaga docente para Banco de Dados. O prof. comunicou que já solicitou uma renovação do edital, mas que até o momento tal solicitação está em trâmite no setor de recursos humanos. O Sr. comunicou que foi criado uma lista de e-mails com os e-mails de todos os alunos do curso e docentes do curso, sendo que esta foi uma solicitação da coordenação do curso para que houvesse maior facilidade na comunicação entre alunos e docentes. Foi destacado que os docentes, por padrão, não recebem os e-mails da lista, estando autorizados apenas ao envio de mensagens. Caso desejem receber e-mails da lista o docente deve se comunicar com o Sr. Rubens.

Ordem do Dia: II) APROVAÇÃO DAS FICHAS DE CARACTERIZAÇÃO: (I) A prof. apresentou as fichas de caracterização que seriam analisadas, esclarecendo que todas elas foram criadas considerando fielmente o Projeto Político Pedagógico do curso. As fichas contemplam as disciplinas obrigatórias do curso que ainda não foram oferecidas e as disciplinas optativas que poderão ser oferecidas no primeiro semestre de 2011, sendo respectivamente as obrigatórias: Física para Computação, Algoritmos e Complexidade, Trabalho de Graduação 1, Estágio Supervisionado 1, Trabalho de Graduação 2, Estágio Supervisionado 2 e Seminários de Computação; e as optativas: Tópicos Avançados em Ciência da Computação, Tópicos Avançados em Redes de Computadores e Sistemas Distribuídos, Segurança e Auditoria de Sistemas e Mineração de Dados. Além das fichas foram apresentados os checklists das respectivas disciplinas, que foi uma solicitação da Coordenação Acadêmica para este semestre. **DISCIPLINAS OPTATIVAS:** (I, II) A prof. destacou que haviam sido elaboradas fichas de quatro disciplinas optativas, porém somente três disciplinas serão oferecidas. Desta maneira haverá uma maior flexibilidade para decidir dentro as quatro disciplinas quais seriam as três oferecidas no próximo semestre. **DISCIPLINAS DE FÍSICA PARA COMPUTAÇÃO E ALGORITMOS E COMPLEXIDADE:** (II, III) A prof. também colocou para análise as fichas de caracterização das disciplinas de Física para Computação e Algoritmos e

Complexidade, explicando que estas disciplinas foram readequadas segundo discussões anteriores sobre a oferta e demanda das mesmas. Juntamente com as fichas de caracterização foi colocada uma ofício justificando a alteração das disciplinas. **AValiação DAS FICHAS:** (IV) A prof. colocou as fichas para que fosse avaliadas pelos membros do conselho. O prof. questionou se quando o aluno vai se matricular na disciplina de Estágio ou de Trabalho de Graduação o sistema verifica se o pré-requisito de créditos mínimos cursados é checado. O prof. respondeu que acreditava que esta verificação era realizada. A prof. colocou que no checklist da disciplina de Algoritmos e Complexidade estava que a disciplina era pré-requisito para outras disciplinas do curso, perguntando se realmente havia algum disciplina que utilizava essa como pré-requisito. A prof. consultou o Projeto Político Pedagógico e verificou que existe uma disciplina optativa, Tópicos Avançados em Teoria da Computação que tem como pré-requisito a disciplina de Algoritmos e Complexidade. (IV) As fichas foram aprovadas pelo Conselho. **Inserimento:** Estarão todos de acordo e nada mais havendo a deliberar, leva-se, lê-se, aprova-se e assina-se esta Ata por todos os membros do Conselho do Curso de Bacharelado em Ciência da Computação. Campus Sorocaba, participantes desta reunião que, em 02 (duas) vias, será levada a registro e arquivamento junto à Coordenação do Conselho do Curso de Bacharelado em Ciência da Computação – Sorocaba, ficando ali à disposição para consulta restrita aos professores da UFSCar – Sorocaba. Nada mais.

Prof. Dr.

Presidente desta Reunião

Prof. Dr.

Professor Associado

Prof. Dr.

Professor Adjunto

Profa. Dra

Professora Adjunta

Profa. Dra

Professora Adjunta

Prof. Dr.

Professor Adjunto

Prof. Dr.

Professor Adjunto

Prof. Dr.

Professor Adjunto

Representante Discente - 2008

Prof. Dr.

Professor Adjunto

Representante Discente - 2009

Representante Discente - 2010

dia, realizada; chamada; estado; conselho;

docentes; técnica; administrativo; presidente; dia;

computação; conselho; aprovado; acordo; ficou;

cursadas; conselho; coordenação; computação; presidente;

disciplinas; cursadas; livros; conselho; aprovado;

representante; discente; presidente; secretária; turma;

como os dados são apresentados? - Tópico com maior similaridade com a consulta - MSV
(expansão de espaço?)

Análise dos Resultados

