



UNIVERSIDADE FEDERAL DE UBERLÂNDIA

Faculdade de Computação

Av. João Naves de Ávila, nº 2121, Bloco 1A - Bairro Santa Mônica, Uberlândia-MG, CEP 38400-902

Telefone: (34) 3239-4144 - <http://www.portal.facom.ufu.br/> facom@ufu.br



PLANO DE ENSINO

1. IDENTIFICAÇÃO

Componente Curricular:	Organização e Recuperação de Informação						
Unidade Ofertante:	Faculdade de Computação						
Código:	GSI521	Período/Série:	5o		Turma:	S	
Carga Horária:				Natureza:			
Teórica:	30	Prática:	30	Total:	60	Obrigatória(X)	Optativa: ()
Professor(A):	Carlos Cesar Mansur Tuma				Ano/Semestre:	2022/2	
Observações:	Observações: a) E-mail institucional do docente: carlos.tuma@ufu.br ; b) Disciplina ofertada de forma presencial cuja aprovação e execução seguem em conformidade com a Resolução CONGRAD nº 73/2022 que aprova os calendários acadêmicos para 2022/1, 2022/2, 2023/1 e 2023/2. c) Ao se matricular na disciplina, o discente declara-se ciente das normas estabelecidas nesse plano de ensino e nas resoluções supracitadas.						

2. EMENTA

Conceitos de documento, palavra e termo. Indexação de documentos: extração de termos, *stopwords*, *stemming*, criação de índices. Thesauri. Modelos de processamento de consultas. Avaliação de Sistemas de Recuperação de Informação (RI). RI em Documentos semiestruturados, multimídia e documentos na Web. Extração da informação. Classificação de documentos. Redução de dimensionalidade.

3. JUSTIFICATIVA

Com a expansão constante das Tecnologias da Informação e Comunicação (TICs), a forma como os dados são armazenados e como as informações são recuperadas passou a ter um papel importante nas organizações, sendo inclusive um diferencial competitivo. Considerando que as informações estão armazenadas em diversos formatos e suportes, inclusive pela Web, é fundamental que o profissional esteja preparado para propor soluções eficazes para sua organização e recuperação.

4. OBJETIVO

Objetivo Geral:

Ao final do curso o aluno estará apto a propor soluções para o problema de recuperar informações nos documentos da coleção que melhor atendam às necessidades do usuário, a partir de uma consulta formulada pelo próprio usuário e de uma coleção de documentos não estruturados ou semi estruturados, como por exemplo uma base de dados em XML, de textos ou de imagens digitais. Além disso, o aluno será apto a avaliar a qualidade de sistemas de recuperação de informação por meio de medidas propostas na literatura.

Objetivos Específicos:

Extração de termos; construção de índices; compressão de índices; modelo vetorial; expansão de consultas; modelo probabilístico; avaliação de sistemas de ORI.

5. PROGRAMA

1. Introdução à Recuperação de Informação e modelo booleano;
2. Dicionário e lista de postings: conceitos de documento, palavra e termo;

3. Indexação de documentos: termos, stopwords, stemming, Thesauri;
4. Compressão de índices;
5. Peso de termos;
6. Modelo Vetorial;
7. Avaliação de sistemas de recuperação de informação;
8. Realimentação de relevantes e expansão de consultas;
9. Recuperação em documentos semiestruturados (XML);
10. Modelo Probabilístico;
11. Tópicos especiais em RI;
 - 11.1 Classificação e agrupamento de documentos;
 - 11.2 Redução de dimensionalidade;
 - 11.3 Web (busca, crawling, indexação, análise de links);
 - 11.4 Extração de informação;
 - 11.5 Introdução à recuperação de imagens baseada em conteúdo.

6. METODOLOGIA

O curso será desenvolvido através de atividades presenciais.

- a) Art. 2º da Resolução CONGRAD Nº 56/2022 “A carga horária dos componentes curriculares será ministrada durante o período previsto neste Calendário e, em caso de carga horária faltante, esta deverá ser complementada por atividades extras, previstas no plano de ensino do componente curricular”. Além disso, no caso da utilização de TDE nos termos dos Art. 2º e 3º da Resolução CONGRAD Nº 73/2022, essas atividades devem estar previstas no plano de ensino. Assim, é necessário prever a totalidade da carga horária da disciplina no plano de ensino com atividades presenciais, de acordo com o horário aprovado pelo Colegiado de Curso, inclusive a complementação necessária e TDE (por exemplo, com vistas das avaliações).
- b) Para a realização plena das atividades presenciais, deverão ser cumpridas a Portaria REITO Nº 287/2022, que dispõe sobre os procedimentos a serem adotados para comprovação do esquema vacinal contra a covid-19 e do acesso aos campi e aos espaços físicos no âmbito da Universidade Federal de Uberlândia para frequentar os espaços internos da UFU, as normas propostas pelo Protocolo de Biossegurança da UFU e o Protocolo Interno de Biossegurança da FACOM;
- c) Atendimento ao discente: devem ser indicados dia(s) da semana, horário e local para o atendimento, e NÃO deve ser condicionado ao agendamento dos discentes. Os horários de atendimento para disciplinas/turmas diferentes podem ser sobrepostos, desde que o total de horas de atendimento do docente atenda à exigência de pelo menos uma hora por disciplina/turma.

O Art. 2º da Resolução CONGRAD No 73/2022 diz “Ficam estabelecidos 100 (cem) dias letivos para o cumprimento do trabalho acadêmico efetivo em cada semestre letivo, com a seguinte distribuição:

I – No mínimo 90 (noventa) dias destinados a aulas, conforme previsto no inciso I do art. 2º da Resolução CES/CNE/MEC Nº 3, de 2 de julho de 2007; e

II – Até 10 (dez) dias destinados a outras atividades acadêmicas, tais como seminários, atividades práticas supervisionadas, atividades em biblioteca, iniciação científica, trabalhos individuais ou em grupo, atividades de recuperação, atividades de extensão, vistas de provas, entre outras, conforme previsto no inciso II do art. 2º da Resolução CES/CNE/MEC Nº 3, de 2 de julho de 2007.

Parágrafo único. Cabe ao professor prever a totalização da carga-horária de cada componente curricular a ser ministrado, no Plano de Ensino, de forma a atender aos incisos I e II do art. 2º da Resolução CES/CNE/MEC nº 3, de 2 de julho de 2007, conforme orientações do Colegiado de Curso.”

O Art. 3º da Resolução CONGRAD No 73/2022 diz “Os professores poderão fazer uso do Trabalho Discente Efetivo - TDE para complementar a carga-horária de componentes curriculares, se necessário.

Parágrafo único. Trabalho Discente Efetivo - TDE corresponde a atividades propostas e orientadas pelos professores, previstas nos Planos de Ensino e realizadas pelos estudantes de forma individual ou em grupo, em horário que for conveniente aos estudantes, respeitando os prazos estabelecidos para a sua conclusão.”

No anexo da Resolução CONGRAD No 73/2022, no calendário do mês de junho de 2023 diz “19 a 29 – Período destinado a outras atividades acadêmicas (10% do semestre letivo)”.

Como no período supracitado teríamos 4 aulas presenciais duplas, foram criados 4 TDEs para cumprir com a carga horária.

Ainda olhando o calendário acadêmico e fazendo a contagem dos dias letivos, considerando apenas os 90%, percebe-se que há uma falta de carga na ordem de 2 horas-aula. Estas foram inseridas como aula de reposição a ser marcada em comum acordo com os alunos.

As atividades presenciais serão aulas expositivas teóricas, exercícios resolvidos e aulas em ambientes de programação, tanto no computador do professor como no do aluno, seminários, prova oral e questionário no Moodle.

As atividades presenciais extras serão aulas de reposição a serem marcadas no primeiro dia de aula e serão voltadas para a resolução de dúvidas sobre cada trabalho.

6.1 - Atividades presenciais: 62 horas-aula.

Horários das atividades presenciais:

Nas 2^{as} das 8:50 às 10:40 e nas 4^{as} das 10:40 às 12:20: aulas de conteúdo, exercícios, entregas de trabalhos e avaliações.

6.2 - Atendimento ao discente:

Nas 2^{as} e 3^{as}, das 10:40 às 11:40 e das 15:00 às 16:00 e quando agendada por e-mail em comum acordo entre professor e discente.

Local para o atendimento: Sala 1A419.

6.3 - Atividades presenciais extras: 2 horas-aula.

A data e o horário da atividade abaixo serão acordados com os alunos no primeiro dia de aula. Data a ser sugerida: entre 04 e 17/04.

Reposição - Apresentação dos resultados: título, resumo e quicklinks: 2 horas-aula.

6.4 - Tecnologias Digitais de Informação e Comunicação (TDIC) que serão utilizadas:

Ambiente de Virtual de Aprendizagem: Moodle

Ambiente de web conferência: Webex da Cisco

Ambiente de programação colaborativa: GitHub

Será usado a linguagem Python versão 3.9.

Os links para instalação dos softwares serão enviados para os alunos matriculados antes do início das aulas.

E-mail e WhatsApp também serão usados como apoio na comunicação e envio de material.

6.5 - Referências bibliográficas: conforme deliberado pelo Colegiado e NDE do BSI-MC, além das referências bibliográficas indicadas no PPC para esta disciplinas, novos materiais poderão ser indicados no início das aulas.

Todo o material produzido e divulgado pelo docente, como vídeos, textos, arquivos de voz etc., está protegido pela Lei de Direitos Autorais (Lei nº 9.610, de 19 de fevereiro de 1998), pela qual fica vetado o uso indevido e a reprodução não autorizada de material autoral por terceiros cabendo, aos responsáveis pela reprodução ou uso indevido do material de autoria dos docentes, as sanções administrativas e às dispostas na Lei de Direitos Autorais.

O discente ao se matricular nesta disciplina concorda com o estabelecido neste plano de ensino.

Para a realização plena das atividades presenciais, deverão ser cumpridas a Portaria REITO Nº 287/2022, que dispõe sobre os procedimentos a serem adotados para comprovação do esquema vacinal contra a covid-19 e do acesso aos campi e aos espaços físicos no âmbito da Universidade Federal de Uberlândia para frequentar os espaços internos da UFU, as

normas propostas pelo Protocolo de Biossegurança da UFU e o Protocolo Interno de Biossegurança da FACOM, e a Resolução CONGRAD nº 58/2022 que dispõe sobre o formato da oferta dos componentes curriculares em 2022/1 e estabelece as regras acerca do uso de máscaras.

7. AVALIAÇÃO

O aluno estará aprovado se a sua frequência for superior ou igual a 75% e a nota final for maior ou igual a 60 pontos. Se a frequência for inferior a 75% será reprovação automática.

A avaliação de desempenho se dará por participações em aula presencial e entrega de atividades avaliativas, individuais ou em grupo, com as seguintes pontuações:

- * participação ativa em aula presencial sem atividade avaliativa - 10 pontos extras, um ponto por aula;
- * entrega de 4 atividades avaliativas (provas) presenciais individuais – 60 pontos, 15 pontos por prova. Estas provas serão na forma de questionário sem consulta;
- * entrega e apresentação de atividades avaliativas presenciais em grupo – 40 pontos. Estas atividades serão o desenvolvimento e entrega de 4 trabalhos, todos valendo 10 pontos cada. Todos os trabalhos deverão ser conter relatório (entregue no Moodle) e seminário presencial.

A nota final é dada pela soma de todas as avaliações dadas, não ultrapassando 100.

Caso o aluno não obtenha o rendimento mínimo para aprovação e tenha frequência mínima de 75% (setenta e cinco por cento), poderá fazer uma das duas atividades avaliativas de recuperação de aprendizagem a seguir:

- Trabalho adicional proposto pelo professor valendo 10 pontos.

A data de entrega e apresentação do trabalho deve ser anterior em pelo menos 1 dia à data do prazo final para o lançamento de notas do período letivo de 2022/1.

A nota obtida no trabalho é somada à nota final.

Caso a nova nota final seja inferior a 60 o aluno está reprovado, caso seja maior ou igual a 60 o aluno está aprovado e a nota é fixada em 60.

- Prova com o total da disciplina valendo 100 pontos.

A data de aplicação da prova deve ser anterior em pelo menos 1 dia à data do prazo final para o lançamento de notas do período letivo de 2022/1.

A nota desta prova será somada a nota final e este total será dividido por 2. Caso esta média seja inferior a 60 o aluno está reprovado mantendo a nota final anterior, caso seja maior ou igual a 60 o aluno está aprovado e a nota é fixada em 60.

Datas e horários da avaliação:

27/03 – 1ª prova, nos horários das aulas presenciais;

24/04 – 2ª prova, nos horários das aulas presenciais;

22/05 – 3ª prova, nos horários das aulas presenciais;

13/06 – 4ª prova, nos horários das aulas presenciais;

20/03 – Entrega do 1º trabalho, nos horários das aulas presenciais;

17/04 – Entrega do 2º trabalho, nos horários das aulas presenciais;

15/05 – Entrega do 3º trabalho, nos horários das aulas presenciais;

06/06 – Entrega do 4º trabalho, nos horários das aulas presenciais.

Crítérios para a realização e correção das avaliações: exatidão das respostas, aparência dos relatórios, corretude dos programas, participação ativa nos seminários.

Validação da assiduidade dos discentes: A frequência (72 presenças) será medida pela:

* (62 horas-aula) pela presença nas aulas presenciais, 2 horas-aula por horário duplo;

* (8 horas-aula) pelo uso de TDEs, 2 horas-aula por TDE;

Datas dos TDEs:

05/03 TDE 1 – Análise morfológica

03/04 TDE 2 – Revisão de Álgebra Linear

15/05 TDE 3 – Aumentando a eficiência no Modelo Vetorial

11/06 TDE 4 – Avaliação de RIs

* (2 horas-aula) pela presença nas aulas presenciais extras: 2 horas-aula por cada reposição em horário duplo.

Especificação das formas de envio das avaliações pelos discentes: A entrega de atividades não avaliativas e relatórios dos trabalhos ocorrerão somente pelo Moodle, nos prazos estipulados na descrição das tarefas.

8. BIBLIOGRAFIA

Básica

BAEZA-YATES, R.; RIBEIRO NETO, B. **Modern information retrieval**. 2. ed. São Paulo AddisonWelsey, 2011.

MANNING, C.; RAGHAVAN, R; SCHÜTZE, H. **An introduction to information retrieval**. Cambridge: Cambridge University Press, 2009.

ZIVIANI, N. **Projeto de Algoritmos: com implementações em Pascal e C**. 3. ed. São Paulo: Cengage Learning, 2011.

Complementar

CRESTANI, F.; PASI, G. **Soft computing in information retrieval: techniques and applications**. New York: Springer Verlag, 2010.

CROFT, B.; METZLER, D.; STROHMAN, T. **Search engines: information retrieval in practice**. São Paulo: Addison Wesley, 2009.

ELMASRI, R.; NAVATHE, S. **Sistemas de Banco de Dados**. 6. ed. São Paulo: Addison-Wesley, 2011.

FRAKES, W. B.; BAEZA-YATES, R. **Information retrieval & data structures**. New Jersey: Prentice Hall, 1992.

MOENS, M. R **Information extraction: algorithms and prospects in a retrieval context**. New York: Springer Verlag, 2006.

REFERÊNCIAS ONLINE COMO MATERIAL DE APOIO

Introduction to Information Retrieval <https://nlp.stanford.edu/IR-book/>

Outros materiais: <https://nlp.stanford.edu/IR-book/information-retrieval.html>

9. APROVAÇÃO

Aprovado em reunião do Colegiado realizada em: ____/____/____

Coordenação do Curso de Graduação: _____