


UNIVERSIDADE FEDERAL DE UBERLÂNDIA
Faculdade de Computação

Av. João Naves de Ávila, 2121, Bloco 1B - Bairro Santa Mônica, Uberlândia-MG, CEP 38400-902
 Telefone: (34) 3239-4393 - www.facom.ufu.br - facom@ufu.br


PLANO DE ENSINO
1. IDENTIFICAÇÃO

Componente Curricular:	Organização e Recuperação da Informação								
Unidade Ofertante:	Faculdade de Computação								
Código:	FACOM32605		Período/Série:		6º		Turma:	S	
Carga Horária:						Natureza:			
Teórica:	30	Prática:	30	Total:	60	Obrigat ória:	(X)	Optativ a:	()
Professor(A):	Renato Aparecido Pimentel da Silva					Ano/Semestre:		2026/2	
Observações:	Plano conforme Res. CONGRAD n.º 158/2025								

2. EMENTA

Conceitos de documento, palavra e termo. Indexação de documentos: extração de termos, *stopwords*, *stemming*, criação de índices. Thesauri. Modelos clássicos de processamento de consultas. Avaliação de Sistemas de RI. RI em documentos semiestruturados. Classificação de documentos. RI na Web.

3. JUSTIFICATIVA

Uma grande motivação da disciplina é entender os diferentes conceitos e tecnologias associadas à construção de máquinas de busca modernas como o Google e o Bing.

4. OBJETIVO

Objetivo Geral:
Ao final do curso, o aluno estará apto a propor soluções para o problema de recuperar informações (RI) nos documentos da coleção que melhor atendam às necessidades do usuário, a partir de uma consulta formulada pelo próprio e de uma coleção de documentos não estruturados ou semiestruturados, como, por exemplo, uma base de dados em XML ou de textos. 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:
Dominar os conceitos dos três modelos clássicos de RI (booleano, vetorial e probabilístico), assim como as técnicas frequentemente utilizadas para a construção de uma máquina de busca, como indexação de documentos utilizando índices invertidos e o pré-processamento de documentos (análise léxica, eliminação de stopwords, stemming e seleção de palavras-chave).

5. PROGRAMA

1. Introdução à Recuperação da Informação
2. Modelos de RI: caracterização de um modelo, taxonomia de modelos de RI
3. Modelo Booleano
4. Ponderação de termos: TF, IDF, TF-IDF

5. Modelo Vetorial
6. Modelo Probabilístico
7. Modelos alternativos: BM25
8. Indexação de documentos: termos, *stopwords*, *stemming*, Thesauri
9. Avaliação de sistemas de recuperação de informação
10. Realimentação de relevantes e expansão de consultas
11. Recuperação em documentos semiestruturados (XML)
12. Classificação de documentos
13. Agrupamento de documentos
14. Web: busca, *crawling*, indexação, análise de links
15. PageRank e Aprendizado de Ranking

6. METODOLOGIA

O curso será ministrado por meio de aulas expositivas, teóricas, sobre o tema, bem como aulas práticas em laboratório, em que serão propostas atividades avaliativas presenciais e trabalhos práticos assíncronos. As aulas ocorrem às segundas-feiras, de 19:00 até 20:40, e às quartas-feiras, de 20:50 até 22:30. Para a exposição, serão utilizados *slides*, disponibilizados em meio virtual, em conjunto com a exposição oral do professor. Sempre que necessário, as explicações e demonstrações que complementam o apresentado em tais transparências serão feitas no quadro da sala. Algumas aulas teóricas eventualmente ocorrerão em laboratório, onde também serão realizadas as atividades avaliativas. Serão, ao todo, 60 horas-aula presenciais. As horas-aula faltantes para cumprimento dos 100 dias letivos (inciso II do Art. 2º da Res. CONGRAD nº 158/2025), bem como da carga horária prevista de 72 horas-aula, serão contabilizadas por meio da proposição de listas de exercícios e implementação dos trabalhos práticos, conforme Art. 3º da Resolução.

O conteúdo programático da disciplina será desenvolvido com auxílio do ambiente virtual de aprendizagem MS Teams – equipe FACOM32605 - 2026/2 - [grupo.ufu.br], código de equipe **43703i4**.

6.1. CRONOGRAMA DE ATIVIDADES PREVISTO

Semana	Módulo	Atividades Presenciais	Carga Horária Presencial	Data e Horário de Atividades Presenciais	Atividades Complementares	Carga Horária das Atividades Complementares
01/09/2026	Início Semestre	-	-	-	-	-
1	Apresentação do professor e da disciplina	Aulas presenciais expositivas.	2 horas-aula	02/09/2026 - 20:50		
2	Tópico 1 - Introdução	Aulas presenciais expositivas. Discussões e debates sobre o módulo.	2 horas-aula	09/09/2026 - 20:50		
3	Tópico 1 - Introdução Tópico 2 - Caracterização de modelos e	Aulas presenciais expositivas. Discussões e debates	4 horas-aula	14/09/2026 - 19:00 16/09/2026 - 20:50		

	modelo booleano	sobre o módulo.				
4	Tópico 2 - Caracterização de modelos e modelo booleano TP1 - Processamento de arquivos, geração de vocabulário e bag of words Tópico 3 - Ponderação de termos	Aulas presenciais expositivas. Discussões e debates sobre o módulo. Atividade avaliativa em grupo assíncrona.	6 horas-aula	21/09/2026 - 19:00 22/09/2026 - 19:00 23/09/2026 - 20:50		
5	Tópico 3 - Ponderação de termos Atendimento e dúvidas sobre o TP1 e QP1	Aulas presenciais expositivas. Discussões e debates sobre o módulo.	4 horas-aula	28/09/2026 - 19:00 30/09/2026 - 20:50	Conclusão assíncrona do TP1	3 horas-aula
6	QP1 - Introdução à RI; conceitos básicos, bag-of-words e modelo booleano Tópico 4 - Modelo vetorial	Atividade avaliativa individual presencial. Aulas presenciais expositivas. Discussões e debates sobre o módulo.	4 horas-aula	05/10/2026 - 19:00 07/10/2026 - 20:50		
7	Tópico 4 - Modelo vetorial TP2 - Leitura de múltiplos arquivos e TF-IDF	Aulas presenciais expositivas. Discussões e debates sobre o módulo. Atividade avaliativa em grupo assíncrona.	4 horas-aula	14/10/2026 - 20:50 15/10/2026 - 19:00		
8	Tópico 5 - Modelo probabilístico	Aulas presenciais expositivas. Discussões e debates sobre o módulo.	4 horas-aula	19/10/2026 - 19:00 21/10/2026 - 20:50	Conclusão assíncrona do TP2	3 horas-aula
9	Tópico 6 - Pré-processamento	Aulas presenciais expositivas.	2 horas-aula	26/10/2026 - 19:00		

		Discussões e debates sobre o módulo.				
10	Tópico 7 - Avaliação da Recuperação	Aulas presenciais expositivas. Discussões e debates sobre o módulo.	2 horas-aula	04/11/2026 - 20:50		
11	TP3 - Modelo Vetorial Tópico 7 - Avaliação da Recuperação Atendimento e dúvidas sobre o TP3 e QP2	Atividade avaliativa em grupo assíncrona. Aulas presenciais expositivas. Discussões e debates sobre o módulo.	6 horas-aula	09/11/2026 - 19:00 11/11/2026 - 20:50 14/11/2026 - 19:00		
12	QP2 - Ponderação de termos; modelos vetorial e probabilístico; comparativo entre modelos Tópico 8 - Realimentação e expansão	Aulas presenciais expositivas. Discussões e debates sobre o módulo.	4 horas-aula	16/11/2026 - 19:00 18/11/2026 - 20:50	Conclusão assíncrona do TP3	3 horas-aula
13	Tópico 8 - Realimentação e expansão; Tópico 9 - Análise de links	Aulas presenciais expositivas. Discussões e debates sobre o módulo.	4 horas-aula	23/11/2026 - 19:00 25/11/2026 - 20:50		
14	TP4 - Classificação de documentos e análise de sentimentos Tópico 9 - Análise de links	Atividade avaliativa em grupo assíncrona. Aulas presenciais expositivas. Discussões e debates sobre o módulo.	4 horas-aula	30/11/2026 - 19:00 02/12/2026 - 20:50		
15	QP3 - avaliação da recuperação; realimentação	Atividade avaliativa individual presencial.	4 horas-aula	07/12/2026 - 19:00 09/12/2026 - 20:50	Conclusão assíncrona do TP4	3 horas-aula

	e expansão de buscas; análise de links Atendimento e dúvidas sobre o TP4 e QP3	Discussões e debates sobre o módulo.				
16	Avaliação de recuperação (prova escrita) Vista da prova substitutiva e encerramento	Atividade avaliativa individual presencial.	4 horas-aula	14/12/2026 - 19:00 16/12/2026 - 20:50		
19/12/2026	Término Semestre	Total de horas-aula presenciais:	60 horas-aula		Total de horas-aula de atividades complementares:	12 horas-aula
Carga Horária Total (presencial + atividades extraclasse):						72 horas-aula

Atendimento ao discente: o atendimento aos alunos ocorrerá semanalmente às quartas-feiras, entre 15:00 e 16:40, e às quintas-feiras, entre 19:00 e 20:40. É necessário agendamento prévio por e-mail, rpimentel@ufu.br, com antecedência de pelo menos um dia, ou pelo fone 32394498, ou ainda por meio do MS Teams. O atendimento será realizado na sala do professor, 1B139.

7. AVALIAÇÃO

A avaliação será feita por meio de três questionários presenciais (quiz, via Teams) – QP, de múltipla escolha sobre o tema – e quatro trabalhos práticos (TP) em grupos, com início de desenvolvimento em sala de aula e posterior conclusão de forma assíncrona. Todas as atividades têm valor de 100 pontos. A nota final será dada pela fórmula $NF = (50 \cdot QP + 50 \cdot TP) / 100$, em que QP e TP são as médias aritméticas simples dos questionários e dos trabalhos, respectivamente. De acordo com o Art. 3º da Res. CONGRAD n.º 158/2025, as conclusões em caráter assíncrono dos trabalhos práticos corresponderão às atividades acadêmicas para complementação da carga horária faltante.

7.1. ATIVIDADE AVALIATIVA DE RECUPERAÇÃO

Conforme o Art. 141 das Normas de Graduação (Res. CONDIR n.º 46/2022), haverá uma avaliação de recuperação de aprendizagem, escrita, dissertativa, que terá valor de 100 pontos. A prova substitutiva abrangerá todo o conteúdo visto no semestre. Segundo o Art. 141, somente fará jus ao direito de realizar a avaliação de recuperação substitutiva o(a) discente que não obtiver o rendimento mínimo de aprovação (60 pontos) e que possuir no mínimo 75% de frequência na disciplina. O discente que for aprovado por meio da prova substitutiva terá aproveitamento lançado de 60 pontos, independentemente da pontuação obtida na prova substitutiva e de outras atividades que eventualmente sejam consideradas. O aluno será aprovado caso $(NA + NR) / 2 \geq 60$, onde NA é a nota até então obtida (dada por NF acima) e NR é a nota de recuperação. Caso $NR < NA$, desconsidera-se NR.

7.2. CONTROLE DE FREQUÊNCIA

A assiduidade será computada por meio da chamada em sala durante as aulas, sempre ao final de cada encontro diário. Não serão admitidos em sala alunos com atraso superior a vinte minutos do início do horário de aula.

7.3. CRONOGRAMA DAS ATIVIDADES AVALIATIVAS

Nro	Data	Hora	Descrição	Pontos
1	22/09/2026	19:00	TP1	100
2	05/10/2026	19:00	QP1	100
3	15/10/2026	19:00	TP2	100

4	09/11/2026	19:00	TP3	100
5	16/11/2026	19:00	QP2	100
6	30/11/2026	19:00	TP4	100
7	07/12/2026	19:00	QP3	100
Recuperação	14/12/2026	19:00 – 20:40	Avaliação de Recuperação	100
TOTAL:				100

8. BIBLIOGRAFIA

Básica

BAEZA-YATES, R.; RIBEIRO-NETO, B. **Recuperação de Informação**: Conceitos e Tecnologia das Máquinas de Busca. Bookman Editora, 2013.

BÜTTCHER, S.; CLARKE, C.; CORMACK, G. **Information retrieval**: Implementing and evaluating search engines. Cambridge: MIT Press, 2010.

MANNING, C.; RAGHAVAN, P.; SCHÜTZE, H. **Introduction to information retrieval**. England: Cambridge University Press, 2008.

Complementar

BAEZA-YATES, R.; RIBEIRO-NETO, B. **Modern information retrieval**. 2nd ed. New York: Addison-Wesley, 2011.

CHOWDHURY, G.; **Introduction to modern information retrieval**. 2nd ed. London: Facet, 2004.

GROSSMAN, D.; FRIEDER, O.; **Information retrieval**: Algorithms and heuristics. Boston: Kluwer, 1998.

KOWALSKI, G.; **Information retrieval architecture and algorithms**. 2nd ed. London: Facet, 2004.

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

9. APROVAÇÃO

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

Coordenação do Curso de Graduação: _____