

Caracterização da disciplina

Código da disciplina:	MCZA017-13	Nome da disciplina:	Processamento de Linguagem Natural						
Créditos (T-P-E-I):	(4-0-0-4)	Carga horária:	48 horas	Aula prática:	N	Campus:	SA		
Código da turma:	DA1MCZA017-13SA	Turma:	A1	Turno:	D	Quadrimestre:	2	Ano:	2024
Recomendação:	Inteligência Artificial								
Docente responsável:	Alexandre Donizeti Alves (alexandre.donizeti@ufabc.edu.br)								

Alocação da turma

	Segunda	Terça	Quarta	Quinta	Sexta	Sábado
8:00 - 9:00		SEMANAL				
9:00 - 10:00						
10:00 - 11:00		ATENDIMENTO		SEMANAL		
11:00 - 12:00						
12:00 - 13:00						
13:00 - 14:00						
14:00 - 15:00						
15:00 - 16:00						
16:00 - 17:00						
17:00 - 18:00						
18:00 - 19:00						
19:00 - 20:00						
20:00 - 21:00						
21:00 - 22:00						
22:00 - 23:00						

Planejamento da disciplina
Objetivos

Estudar os problemas da geração e compreensão automática de línguas humanas naturais.

Ementa

Introdução ao processamento de linguagem natural. Processamento sintático. Técnicas de análise (*parsing*). Gramáticas. Interpretação semântica. Processamento de discurso. Aplicações.

Conteúdo programático

Aula	Data	Conteúdo	Estratégias didáticas	Avaliação/Observação
01	25/06	Apresentação da Disciplina Introdução ao Processamento de Linguagem Natural	Slides	Perfil da Turma
02	27/06	Expressões Regulares	Slides e Notebooks	
03	02/07	Extração e Pré-processamento de Dados	Notebooks	
04	04/07	Normalização de Textos	Slides e Notebooks	
05	11/07	Modelos de Linguagem com N-gramas	Slides e Notebooks	
06	16/07	Representação de Textos	Slides e Notebooks	
07	18/07	AVALIAÇÃO 01 (20%) [INDIVIDUAL]	PROVA ESCRITA	TODA A MATÉRIA
08	23/07	Introdução à API da OpenAI	Slides e Notebooks	PROJETO PRÁTICO (30%)
09	25/07	Grandes Modelos de Linguagem	Slides e Notebooks	
10	30/07	LangChain	Slides e Notebooks	
11	01/08	LangChain	Slides e Notebooks	
12	06/08	Extração de Informação: Reconhecimento de Entidades Nomeadas e Sumarização de Textos	Slides e Notebooks	
13	08/08	Categorização de Textos: Classificação de Textos e Análise de Sentimentos	Slides e Notebooks	
14	13/08	Categorização de Textos: Similaridade de Textos	Slides e Notebooks	
15	15/08	Recuperação de Informação: Sistemas de Perguntas e Respostas	Slides e Notebooks	

16	22/08	AVALIAÇÃO 02 (30%) [INDIVIDUAL]	PROVA ESCRITA	TODA A MATÉRIA
17	27/08	SEMINÁRIO [PRÁTICA]	Notebooks	(20%) [EM GRUPO]
18	29/08	SEMINÁRIO [PRÁTICA]	Notebooks	
19	03/09	SEMINÁRIO [PRÁTICA]	Notebooks	
20	05/09	SEMINÁRIO [PRÁTICA]	Notebooks	
21	10/09	SEMINÁRIO [PRÁTICA]	Notebooks	
22	12/09	AVALIAÇÃO SUBSTITUTIVA	PROVA ESCRITA	TODA A MATÉRIA
23	17/09	VISTA		
24	19/09	AVALIAÇÃO DE RECUPERAÇÃO	PROVA ESCRITA	TODA A MATÉRIA

Descrição dos instrumentos e critérios de avaliação qualitativa

Detalhamento das estratégias didáticas a serem usadas:

- Disponibilização de conteúdo prático por meio de Cadernos de Atividades (Notebooks/Python), com conteúdo descritivo (textos, figuras, exemplos de código, links para conteúdos etc.). Os cadernos de atividades poderão ser visualizados utilizando a ferramenta **Google Colab**.
- Os slides das aulas serão disponibilizados no Ambiente Virtual de Aprendizagem **Moodle**.

Recursos e Ferramentas:

Para acompanhar a disciplina são necessários os seguintes recursos e ferramentas:

- Computador com acesso à Internet;
 - Navegador Web compatível com o Ambiente Virtual de Aprendizagem **Moodle**. O navegador também deve permitir a execução de notebooks no **Google Colab** a partir do **Google Drive**. A **sugestão** é o utilizar o **Google Chrome**.
 - Linguagem de Programação Python **3.12.4** (<https://www.python.org/downloads>).
- Recomenda-se a instalação da plataforma Anaconda (<https://www.anaconda.com/download>) ou Visual Studio Code (<https://code.visualstudio.com/>).

IMPORTANTE: instalando a plataforma Anaconda, não é necessário instalar o Python separadamente.

Critérios de Avaliação:

A composição do **Conceito Final** durante o Quadrimestre será obtida da seguinte forma:

- **AVALIAÇÃO 01 (Peso 20%*)**: prova escrita e individual
Será realizada no dia **18/07/2024** (quinta-feira).
- **AVALIAÇÃO 02 (Peso 30%*)**: prova escrita e individual
Será realizada no dia **22/08/2024** (quinta-feira).
- **PROJETO PRÁTICO (Peso 30%*)**: em dupla ou em trio
Disponível a partir do dia **25/07/2024** (quinta-feira). Deve ser entregue até o dia **08/09/2024** (domingo).
- **SEMINÁRIO (Peso 20%*)**: em dupla ou em trio
Disponível a partir do dia **23/07/2024** (terça-feira). As apresentações serão nos seguintes dias:
27/08/2024 (terça-feira) **29/08/2024** (quinta-feira) **03/09/2024** (terça-feira)
05/09/2024 (quinta-feira) **10/09/2024** (terça-feira)

Os critérios de avaliação e as datas de entrega serão definidos e disponibilizados (**com ANTECEDÊNCIA**) no Ambiente Virtual de Aprendizagem **Moodle**.

A atribuição do **Conceito Final** seguirá a seguinte conversão:

- A** se **NOTA** $\geq 9,0$
- B** se $7,5 \leq \text{NOTA} < 9,0$
- C** se $6,0 \leq \text{NOTA} < 7,5$
- D** se $5,0 \leq \text{NOTA} < 6,0$
- F** se **NOTA** $< 5,0$ (**REPROVADO**)
- O** se frequência $< 75\%$ (**REPROVAÇÃO POR FALTAS**)

Legenda:

- A**: desempenho excepcional, demonstrando excelente compreensão da disciplina
- B**: bom desempenho, demonstrando boa capacidade de uso dos conceitos da disciplina
- C**: desempenho adequado, demonstrando capacidade de uso dos conceitos da disciplina e capacidade para seguir estudos mais avançados
- D**: aproveitamento mínimo dos conceitos da disciplina com familiaridade parcial do assunto, mas demonstrando deficiências que exigem trabalho adicional para prosseguir em estudos avançados
- F**: reprovado. A disciplina deve ser cursada novamente para a obtenção do crédito

A **reprovação por faltas** (conceito **O**) ocorre caso a frequência seja inferior a 75% (resolução ConsEPE 139). As faltas poderão ser justificadas de acordo com os critérios estabelecidos na resolução ConsEPE 227.

IMPORTANTE: Caso seja verificado a ocorrência de **PLÁGIO** nos PROJETOS/SEMINÁRIOS, **TODOS** estudantes envolvidos serão automaticamente **REPROVADOS** com CONCEITO FINAL **F**. Lembrem-se que a UFABC possui um Código de Conduta (Ética) e o mesmo poderá ser aplicado, caso seja necessário.

Avaliação Substitutiva:

Estarão habilitados para submeter novamente uma AVALIAÇÃO os estudantes que se ausentarem a uma das avaliações do período e contemplados pelo benefício de acordo com a Resolução CONSEPE no. 227, de 23 de abril de 2018. Nesse caso, deverá ser enviada uma justificativa válida ao docente responsável pela disciplina cujo período de ausência compreenda todo o período na qual a atividade perdida esteve disponível para entrega.

Caso o estudante se ausente de mais de uma avaliação do período regular, o conceito da avaliação substitutiva será concedido para UMA ÚNICA avaliação não realizada, privilegiando a de maior peso ponderado. Estudantes que fizeram todas as avaliações NÃO TERÃO DIREITO à avaliação substitutiva.

Data da prova substitutiva: **12/09/2024** (quinta-feira, às 10h).

Avaliação de Recuperação:

Estarão habilitados para a avaliação de recuperação os estudantes que obtiverem conceito final **D** ou **F** na conclusão de todas as atividades e avaliações aplicadas no quadrimestre suplementar, obedecendo as regras indicadas na Resolução CONSEPE no. 182, de 23 de outubro de 2014. Observe que o **Conceito Final** máximo obtido neste caso é **C**:

Pré-Recuperação	Recuperação	Conceito Final
D	A	C
D	B	C
D	C	C
D	D	D
D	F	D
F	A	C
F	B	C
F	C	D
F	D	F
F	F	F

Data da prova de recuperação: **19/09/2024** (quinta-feira, às 8h).

Horário de Atendimento:

Esta disciplina prevê um horário de atendimento extraclasse para atividades de apoio aos estudantes regulares desta turma, conforme disposto na Resolução CONSUNI 183, de 31 de outubro de 2017.

O horário de atendimento semanal terá carga horária total de 2 horas, sendo realizado no seguinte dia e horário:

- **Terça-feira, das 10:00h às 12:00h, sala 513-2.**

Eventuais dúvidas e questionamentos poderão ser feitos em outros horários, desde que agendados com antecedência. Nesse caso, o contato deverá ser feito pelo meu e-mail institucional do docente responsável pela disciplina:

- alexandre.donizeti@ufabc.edu.br (informar o nome e RA, obrigatoriamente, por favor).

Os estudantes também serão atendidos (dúvidas sobre conteúdos, atividades ou outros esclarecimentos), **PREFERENCIALMENTE**, por meio das ferramentas de mensagens disponíveis no ambiente virtual de aprendizagem **Moodle**.

Referências bibliográficas básicas

- [1]. ALLEN, J. **Natural language understanding**. 2a edição. Menlo Park: The Benjamin/Cummings, 1995.
- [2]. BRATKO, I. **Prolog programming for artificial intelligence**. Boston, USA: Addison-Wesley Longman, 1986.
- [3]. GAZDAR G.; MELLISH, C. **Natural language processing in PROLOG: an introduction to computational linguistics**. Wokingham, UK; Reading, USA: Addison-Wesley, 1989.

Referências bibliográficas complementares

- [1]. AMBLE, T. **Logic programming and knowledge engineering**. Wokingham, UK: Addison-Wesley, 1987.
- [2]. CHARNIAK, E. **Statistical Language Learning**. Cambridge, UK: MIT Press, 1996.
- [3]. GARSIDE, R. et al. **Corpus annotation: linguistic information from computer text corpora**. Reading, USA: Addison-Wesley, 1997.
- [4]. STERLING, L.; SHAPIRO, E. **The art of Prolog: advanced programming techniques**. Cambridge, UK: MIT Press, 1994.
- [5]. GRISHMAN, R. **Computational linguistics**. Cambridge, UK: Cambridge University Press, 1986.