

Caracterização da disciplina

Código da disciplina:	MCZA017-13	Nome da disciplina:	Processamento de Linguagem Natural						
Créditos (T-P-I):	(4-0-4)	Carga horária:	48 horas	Aula prática:	N	Campus:	SA		
Código da turma:	DA1MCZA017-13SA	Turma:	A1	Turno:	Diurno	Quadrimestre:	1	Ano:	2023
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		Semanal				
10:00 - 11:00				Semanal		
11:00 - 12:00				Semanal		
12:00 - 13:00						
13:00 - 14:00		Atendimento PRESENCIAL		Atendimento PRESENCIAL		
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
01	07/02	Apresentação da Disciplina Recursos e Ferramentas	Slides	
02	09/02	Introdução ao Processamento de Linguagem Natural	Slides	
03	14/02	Expressões Regulares	Slides e Notebooks	
04	16/02	Extração e Pré-processamento de Dados	Slides e Notebooks	
05	23/02	Normalização de Textos	Slides e Notebooks	
06	28/02	Modelo de Linguagem com N-gramas	Slides e Notebooks	
07	02/03	Representação de Textos	Slides e Notebooks	
08	07/03	Representação de Textos	Slides e Notebooks	
09	09/03	AVALIAÇÃO (40%)	PROVA ESCRITA	INDIVIDUAL
10	14/03	Chatbots	Slides	PROJETO PRÁTICO I (20%)
11	16/03	Chatbots	Slides	
12	21/03	Classificação de Textos	Slides e Notebooks	PROJETO PRÁTICO II (40%)
13	23/03	Análise de Sentimentos	Slides e Notebooks	
14	28/03	PROJETO NA PRÁTICA	Slides e Notebooks	
15	30/03	VISTA DE PROVA		
16	04/04	Chatbots com Python	Slides e Notebooks	
17	06/04	Modelagem de Tópicos	Slides e Notebooks	

18	11/04	Reconhecimento de Entidades Nomeadas	Slides e Notebooks	
19	13/04	PROJETO NA PRÁTICA	Slides e Notebooks	
20	18/04	Similaridade de Textos	Slides e Notebooks	
21	20/04	Sumarização de Textos	Slides e Notebooks	
22	25/04	Sistemas de Perguntas e Respostas	Slides e Notebooks	
23	27/04	PROVA SUBSTITUTIVA	TODA A MATÉRIA	INDIVIDUAL
24	04/05	PROVA DE RECUPERAÇÃO	TODA A MATÉRIA	INDIVIDUAL

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 usando 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 (com Webcam, caixa de som ou fone de ouvido);
- Navegador Web compatível com o **Google Meet** e com o Ambiente Virtual de Aprendizagem **Moodle**. O navegador também deve permitir a execução de notebooks no **Google Colab** a partir do **Google Drive**. Uma **sugestão** é o usar o **Google Chrome**.
- Linguagem de Programação Python 3.11.1 (<https://www.python.org/download>).

Recomenda-se a instalação da plataforma Anaconda (<https://www.anaconda.com/products/individual>), disponível para:

- Windows [64 bits] (https://repo.anaconda.com/archive/Anaconda3-2022.10-Windows-x86_64.exe)
- Linux [64 bits] (https://repo.anaconda.com/archive/Anaconda3-2022.10-Linux-x86_64.sh)
- MacOS [64 bits] (https://repo.anaconda.com/archive/Anaconda3-2022.10-MacOSX-x86_64.pkg).

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 (Peso 40%):** prova escrita e individual
Será realizada no dia **09/03/2023**
- **PROJETO PRÁTICO I (Peso 20%):** individual, em dupla ou em trio
Disponível a partir do dia **14/03/2023**. Deve ser entregue até o dia **09/04/2023**.
- **PROJETO PRÁTICO II (Peso 40%):** individual, em dupla ou em trio
Disponível a partir do dia **21/03/2023**. Deve ser entregue até o dia **30/04/2023**.

Os critérios de avaliação dos Projetos/Atividades 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. Como não haverá lista, a presença será definida de acordo com as atividades entregues pelos estudantes. Cada atividade não entregue resultará em 2 (duas) faltas.

IMPORTANTE: Caso seja verificado a ocorrência de **PLÁGIO** nas ATIVIDADES/PROJETOS, **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 ATIVIDADE/PROJETO os estudantes que se ausentarem a uma das Atividades/Projetos 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: 27/04/2023

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: 04/05/2023

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 nos seguintes dias e horários:

- Terça-feira, das 13:00h às 14:00h, sala 513-2.
- Quinta-feira, das 13:00h às 14:00h, 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.