

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:	NAMCZA017-13SA			Turma:	A	Turno:	N	Quadrimestre:		3	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						
9:00 - 10:00						
10:00 - 11:00						
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			ATENDIMENTO PRESENCIAL			
18:00 - 19:00						
19:00 - 20:00			SEM ANAL			
20:00 - 21:00						
21:00 - 22:00	SEM ANAL					
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	18/09 SEG	Apresentação da Disciplina	Slides	ATIVIDADE 01 Prof. Jesús P. Mena-Chalco Prof. Jesús P. Mena-Chalco
02	20/09 QUA	Introdução ao Processamento de Linguagem Natural	Slides	
03	25/09 SEG	Uso do ChatGPT para Ensino e Pesquisa [Aplicações]	Slides	
04	27/09 QUA	Projetos de Pesquisa que usam PLN como ferramenta [Aplicações]	Slides	
05	02/10 SEG	Recursos e Ferramentas [Curadoria] Expressões Regulares [Sintaxe] Introdução ao Google Colab [EXTRA]	Slides	ATIVIDADE 02 PROJETO PRÁTICO (30%)
06	04/10 QUA	Expressões Regulares [Prática] Extração e Pré-processamento de Dados	Notebooks	
07	09/10 SEG	Normalização de Textos	Slides e Notebooks	
08	11/10 QUA	Modelos de Linguagem com N-gramas	Slides e Notebooks	
09	16/10 SEG	Representação de Textos	Slides e Notebooks	ATIVIDADE 03 SEMINÁRIO (20%)
10	18/10 QUA	Introdução à API da OpenAI [Aplicações]	Slides e Notebooks	
11	23/10 SEG	Grandes Modelos de Linguagem	Slides e Notebooks	Prof. Raphael Camargo
12	25/10 QUA	Inteligência Artificial Generativa [Aplicações]	Slides	
13	30/10 SEG	Chatbots [Aplicações]	Slides	IBM Watson Assistant
14	01/11 QUA	Extração de Informação: Reconhecimento de Entidades Nomeadas, Extração de Frases-chave, Sumarização de Textos,	Slides e Notebooks	ATIVIDADE 04

15	06/11 SEG	Detecção de Idiomas e Tradução [Sintaxe] Categorização de Textos: Classificação de Textos e Análise de Sentimentos [Semântica]	Slides e Notebooks	ATIVIDADE 05
16	08/11 QUA	Categorização de Textos: Modelagem de Tópicos e Similaridade de Textos [Semântica]	Slides e Notebooks	
17	13/11 SEG	Recuperação de Informação: Sistemas de Perguntas e Respostas [Semântica, Discurso]	Slides e Notebooks	
18	22/11 QUA	AVALIAÇÃO (40%) [INDIVIDUAL]	PROVA ESCRITA	TODA A MATÉRIA
19	27/11 SEG	SEMINÁRIOS [Aplicações]	Slides	TODA A MATÉRIA
20	29/11 QUA	SEMINÁRIOS [Aplicações]	Slides	
21	04/12 SEG	SEMINÁRIOS [Aplicações]	Slides	
22	06/12 QUA	AVALIAÇÃO SUBSTITUTIVA [INDIVIDUAL]	PROVA ESCRITA	TODA A MATÉRIA
23	19/12 TER	VISTA DE AVALIAÇÃO		FERIADO 15/11: reposição na terça-feira, às 19h
24	20/12 QUA	AVALIAÇÃO DE RECUPERAÇÃO [INDIVIDUAL]	PROVA ESCRITA TODA A MATÉRIA	FERIADO 20/11: reposição na quarta-feira, às 21h

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 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 3.11.5 (<https://www.python.org/download>).
- Recomenda-se a instalação da plataforma Anaconda (<https://www.anaconda.com/products/individual>):
- Windows [64 bits] (https://repo.anaconda.com/archive/Anaconda3-2023.07-2-Windows-x86_64.exe)
 - Linux [64 bits] (https://repo.anaconda.com/archive/Anaconda3-2023.07-2-Linux-x86_64.sh)
 - MacOS [64 bits] (https://repo.anaconda.com/archive/Anaconda3-2023.07-2-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 **22/11/2023** (quarta-feira).
- **PROJETO PRÁTICO (Peso 30%*):** individual, em dupla ou em trio
Disponível a partir do dia **16/10/2023** (segunda-feira).
Deve ser entregue até o dia **10/12/2023** (domingo).
- **SEMINÁRIO (Peso 20%*):** individual, em dupla ou em trio
Disponível a partir do dia **23/10/2023** (segunda-feira). As apresentações serão nos seguintes dias:
27/11/2023 (segunda-feira)
29/11/2023 (quarta-feira)
04/12/2023 (segunda-feira)
- **ATIVIDADES (Peso 20%*):** individual, em dupla ou em trio
ATIVIDADE 01: disponível a partir do dia **20/09/2023** (quarta-feira).
Deve ser entregue até o dia **06/10/2023** (sexta-feira).
ATIVIDADE 02: disponível a partir do dia **04/10/2023** (quarta-feira).
Deve ser entregue até o dia **16/10/2023** (segunda-feira).
ATIVIDADE 03: disponível a partir do dia **18/10/2023** (quarta-feira).
Deve ser entregue até o dia **30/10/2023** (segunda-feira).
ATIVIDADE 04: disponível a partir do dia **01/11/2023** (quarta-feira).
Deve ser entregue até o dia **20/11/2023** (segunda-feira).
ATIVIDADE 05: disponível a partir do dia **08/11/2023** (quarta-feira).
Deve ser entregue até o dia **20/11/2023** (segunda-feira).

*** OS PESOS ESTÃO CORRETOS.**

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. 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 **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: **06/12/2023** (quarta-feira, às 19h).

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: **20/12/2023** (quarta-feira, às 21h).

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:

- **Quarta-feira, das 17:00h às 19:00h, 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.