

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:	DAMCZA017-13SA	Turma:	A	Turno:	Matutino	Quadrimestre:	3	Ano:	2021
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		Atendimento		Semanal		
9:00 - 10:00		Atendimento		Semanal		
10:00 - 11:00		Semanal				
11:00 - 12:00		Semanal				
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	Conteúdo	Estratégias didáticas	Avaliação	
01	Apresentação da Disciplina Recursos e Ferramentas [Curadoria]	Slides e Videoaula [SÍNCRONA]		
02	Introdução ao Processamento de Linguagem Natural	Slides e Videoaula [ASSÍNCRONA]		
03	Expressões Regulares	Slides, Videoaula e Notebooks [ASSÍNCRONA]		
04	Extração e Pré-processamento de Dados	Slides, Videoaula e Notebooks [SÍNCRONA]		
05	Normalização de Textos	Slides, Videoaula e Notebooks [ASSÍNCRONA]		
06	Modelo de Linguagem com N-gramas	Slides, Videoaula e Notebooks [ASSÍNCRONA]		
07	Representação de Textos	Slides, Videoaula e Notebooks [ASSÍNCRONA]		
08	ATIVIDADE PRÁTICA 1 (10%)	Individual ou em Dupla		Avaliação (72h)
09	Chatbots com Google Dialogflow	Slides e Videoaula [SÍNCRONA]		
10	Chatbots com Google Dialogflow	Slides e Videoaula [SÍNCRONA]		
11	Chatbots com Google Dialogflow	Slides e Videoaula [SÍNCRONA]		
12	ATIVIDADE PRÁTICA 2 (25%)	Individual ou em Dupla		Avaliação (72h)
13	Classificação de Textos	Slides, Videoaula e Notebooks [ASSÍNCRONA]		

14	Análise de Sentimentos	Slides, Videoaula e Notebooks ASSÍNCRONA]	Avaliação (72h) [PROJETO: 40%] Prova Escrita/Projeto
15	ATIVIDADE PRÁTICA 3 (10%)	Individual ou em Dupla	
16	Chatbots com Python	Slides, Videoaula e Notebooks SÍNCRONA]	
17	Modelagem de Tópicos	Slides, Videoaula e Notebooks ASSÍNCRONA]	
18	Similaridade de Textos	Slides, Videoaula e Notebooks ASSÍNCRONA]	
19	Sumarização de Textos	Slides, Videoaula e Notebooks ASSÍNCRONA]	
20	Reconhecimento de Entidades Nomeadas	Slides, Videoaula e Notebooks ASSÍNCRONA]	
21	Sistemas de Perguntas e Respostas	Slides, Videoaula e Notebooks ASSÍNCRONA]	
22 e 23	ATIVIDADE PRÁTICA 4 (15%)	Individual ou em Dupla	
24	PROVA DE RECUPERAÇÃO	TODA A MATÉRIA	

Descrição dos instrumentos e critérios de avaliação qualitativa

Detalhamento das estratégias didáticas a serem usadas:

- Videoconferências por meio da ferramenta **Google Meet** (<https://meet.google.com/xbv-jpwh-fgg>).
- As videoconferências (atividade SÍNCRONA) serão gravadas e disponibilizadas para acesso ASSÍNCRONO posterior no **YouTube** (<https://youtube.com/playlist?list=PL2W7er93mQ7XSUQi00yvimsUWOphrQ36Y>).
- 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 vídeos a partir do **YouTube** e notebooks a partir do **Google Drive**. No caso dos notebooks deve permitir a execução no **Google Colab**. Uma **sugestão** é o usar o **Google Chrome**.
- Linguagem de Programação Python 3.9.7 (<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/Anaconda3-2021.05-Windows-x86_64.exe)
- Linux [64 bits] (https://repo.anaconda.com/archive/Anaconda3-2021.05-Linux-x86_64.sh)
- MacOS [64 bits] (https://repo.anaconda.com/archive/Anaconda3-2021.05-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 Suplementar será obtida da seguinte forma:

- **ATIVIDADE PRÁTICA 1 (Peso 10%):** individual ou em dupla
07/10/2021 – disponível a partir de 08h. **Deve ser entregue até o dia 12/10/2021.**
- **ATIVIDADE PRÁTICA 2 (Peso 25%):** individual ou em dupla
26/10/2021 – disponível a partir de 08h. **Deve ser entregue até o dia 02/11/2021.**
- **ATIVIDADE PRÁTICA 3 (Peso 10%):** individual ou em dupla
11/11/2021 – disponível a partir de 08h. **Deve ser entregue até o dia 14/11/2021.**
- **ATIVIDADE PRÁTICA 4 (Peso 15%):** individual ou em dupla
07/12/2021 – disponível a partir de 08h. **Deve ser entregue até o dia 08/12/2021.**
- **PROJETO (Peso 40%):** individual ou em dupla.
disponível a partir do dia **16/11/2021**. **Deve ser entregue até o dia 05/12/2021.**

Os critérios de avaliação dos Projetos 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.

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:

13/12/2021 – disponível a partir de 08h. Deve ser entregue até o dia **16/12/2021**.

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 08:00h às 10:00h, via Google Meet (<https://meet.google.com/xbv-jpwh-fgg>).

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.