

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:	NAMCZA017-13SA	Turma:	A	Turno:	Noturno	Quadrimestre:	QS	Ano:	2020
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						
18:00 - 19:00						
19:00 - 20:00			X			
20:00 - 21:00			X			
21:00 - 22:00	X					
22:00 - 23:00	X					

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

Semana	Conteúdo	Estratégias didáticas
01	Aula 01 – Apresentação da Disciplina	Slides e Videoaula
	Aula 02 – Introdução ao Processamento de Linguagem Natural	Slides e Videoaula
02	Aulas 03 e 04 – Expressões Regulares	Slides, Videoaula e Notebooks
03	Aulas 05 e 06 – Normalização de textos	Slides, Videoaula e Notebooks
04	Aulas 07 e 08 – Modelo de linguagem com N-gramas	Slides, Videoaula e Notebooks
05	Aulas 09 e 10 – Classificação de textos	Slides, Videoaula e Notebooks
06	Aulas 11 e 12 – Análise Semântica	Slides, Videoaula e Notebooks
07	Aulas 13 e 14 – Extração de Informação	Slides, Videoaula e Notebooks
08	Aulas 15 e 16 – Chatbots	Slides, Videoaula e Notebooks
09	Aulas 17 – Reconhecimento de Entidades Nomeadas	Slides, Videoaula e Notebooks
	Aula 18 – Sistemas de Perguntas e Respostas	Slides, Videoaula e Notebooks
10	Aula 19 – Sumarização	Slides, Videoaula e Notebooks
	Aula 20 – Modelagem de Tópicos	Slides, Videoaula e Notebooks
11	Aulas 21 e 22 – Análise de Sentimentos	Slides, Videoaula e Notebooks
12	Aula 23 – Avaliação dos Chatbots (Projeto 02)	Videoaula
	Aula 24 – Avaliação dos Trabalhos (Projeto 03)	Videoaula
13	RECUPERAÇÃO	---

Descrição dos instrumentos e critérios de avaliação qualitativa

Detalhamento das estratégias didáticas a serem usadas:

- Videoconferências por meio da plataforma **Google Meet**. As videoconferências (atividade SÍNCRONA) serão gravadas e disponibilizadas para acesso ASSÍNCRONO posterior.
- 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**.
- Videoaulas assíncronas serão disponibilizadas para algumas aulas, principalmente quando for feriado no dia da respectiva aula.
- Os slides das aulas serão disponibilizados no Ambiente Virtual de Aprendizagem, SIGAA ou 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 os ambientes virtuais de aprendizagem (SIGAA e MOODLE). O navegador também deve permitir a execução de vídeos 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 3.8 (<https://www.python.org/download>).

Recomenda-se a instalação da plataforma Anaconda (versão de julho de 2020), disponível em:

- Windows [64 bits] (https://repo.anaconda.com/archive/Anaconda3-2020.07-Windows-x86_64.exe)
- Linux [64 bits] (https://repo.anaconda.com/archive/Anaconda3-2020.07-Linux-x86_64.sh)

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:

- Atividades Semanais Individuais (**Peso 10%**): ao todo serão 10 atividades, todas com prazo mínimo de 1 semana para a entrega.
- Projeto 01 (**Peso 15%**): grupos de 2 a 3 alunos. Este projeto será dividido em 3 partes (A, B e C).
- Projeto 02 - Chatbot (**Peso 25%**): individual ou em dupla. Este projeto será dividido em 2 partes (A e B).
- Projeto 03 (Peso 50%): grupos de 2 a 3 alunos. Este projeto será dividido em 2 partes (A e B)

Os critérios de avaliação das Atividades Semanais Individuais/Projetos e as datas de entrega serão definidos e disponibilizados (**com antecedência**) no ambiente virtual de aprendizagem (SIGAA ou MOODLE). Todas as entregas de atividades serão assíncronas e deverão ser disponibilizadas no ambiente virtual de aprendizagem (SIGAA ou 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:

semana 13 (14/12/2020 – 17/12/2020)

Será disponibilizada uma ATIVIDADE PRÁTICA até às 21h do dia 14/12/2020 (segunda-feira), com o PRAZO DE ENTREGA de 72 horas, ou seja, até o dia 17/12/2020 (quinta-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:

- Quinta-feira, das 19:00h às 21:00h, via Google Meet (**link a ser disponibilizado no SIGAA**)

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 (SIGAA ou 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.