

UFRN/DIMAp

MODELO DE PLANO DE CURSO PARA COMPONENTE CURRICULAR OFERECIDO PELO DIMAp, EXCLUSIVO PARA A RETOMADA DO PERÍODO LETIVO 2020.1 (COMO 2020.6) NO FORMATO REMOTO

Dados do Componente	
Unidade responsável	Departamento de Informática e Matemática Aplicada
Código	DIM0135
Nome	Processamento de Linguagem Natural
Carga horária teórica	45
Carga horária prática	15

Dados do docente*	
Nome: Carlos Augusto Prolo	SIAPE:
Cargo: Professor Efetivo do Ensino Superior	
Unidade de exercício: Departamento de Informática e Matemática Aplicada – DIMAP	
Celular:	email: prolo@dimap.ufrn.br

*No caso de componente curricular a ser ofertado por mais de um docente, o quadro anterior deve ser replicado.

Conteúdo	Em caso de componente curricular já cadastrado, copie a ementa do SIGAA (na aba Ensino > Consulta > Componentes curriculares
	<ol style="list-style-type: none">1. Overview sobre os desafios de PLN.2. Corpos linguísticos anotados para Inglês e Português, aprendizado supervisionado e avaliação da eficácia.3. Módulos, componentes, tarefas de um sistema de PLN: pré-processamento; processamento ao nível das palavras; processamento ao nível das estruturas sintáticas ou de dependências; processamento ao nível semântico, pragmático, de discurso, resolução de correferência.4. Técnicas de Aprendizado de Máquina para Processamento de Linguagem Natural5. Representações distribuídas para palavras (<i>vector space models</i>, <i>word embeddings</i>)6. Aplicações de PLN: os tópicos poderão variar a cada semestre; candidatos típicos são: NER (named-entity recognition), extração de tópicos relevantes e classificação de documentos, análise de sentimento, sistemas de recomendação, Q&A (perguntas e respostas), aconselhadores ortográficos e sintáticos, tradução (semi-) automática, sumarização, inferência textual, reconhecedores de estilo, outros tópicos Recuperação e Extração de Informações – IR/IE

UFRN/DIMAp

Metodologia	<p>Descrição de como a disciplina será desenvolvida, especificando as técnicas de ensino a serem utilizadas.</p> <p>Disciplinas com carga horária prática devem atentar ao disposto no Art. 2º "Os componentes curriculares de natureza prática ou a parte prática de componentes curriculares poderão ser adaptados ao formato remoto, desde que seja elaborado plano de curso específico, para o período letivo 2020.1, aprovado pelo colegiado de curso e apensado ao Projeto Pedagógico de Curso."</p> <p>Sendo assim, quando for o caso, deve ficar claro como a carga horária prática será adaptada ao formato remoto.</p>
	<p>A disciplina consistirá de aulas síncronas para apresentação de conteúdo e comunicação conjunta como explicação das atividades práticas. A combinar com os alunos, as aulas síncronas poderão ser gravadas e disponibilizadas.</p>
	<p>As aulas assíncronas serão usadas especialmente para a elaboração das atividades práticas conforme abaixo.</p>
	<p>No cronograma, usei síncrona e assíncrona para caracterizar aulas síncronas em que posteriormente ficarei a disposição para complementação seja individual ou por grupos.</p> <p>A parte prática da disciplina consiste na execução de trabalhos de computador pelos alunos, utilizando C++, Python, e ferramentas de desenvolvimento de machine learning pra PLN disponíveis na Internet. Isto pode ser feito remotamente nos computadores dos alunos sob supervisão remota do professor sem problema algum utilizando ferramentas ágeis de comunicação escrita e visual explicitadas abaixo.</p>

Procedimentos de avaliação da aprendizagem	<p>Descrição dos instrumentos e critérios a serem utilizados para a verificação da aprendizagem.</p>
	<p>A avaliação será feita através de três trabalhos práticos desenvolvidos em computador, com datas de entrega prováveis: 07/10, 11/11 e 7/12.</p> <p>A avaliação de reposição, se necessária, consistirá de um conjunto de questões para o aluno fazer assincronamente e entregar na data combinada (provavelmente 14/12)</p>

Cronograma e critérios para validação da assiduidade dos discentes	<p>Cronograma detalhado das atividades e dos critérios de validação da assiduidade dos discentes.</p> <p>(Art. 3º §4º "A frequência e a participação dos discentes serão verificadas de acordo com o acompanhamento das atividades propostas, conforme plano de curso.")</p>
	<ul style="list-style-type: none"> • Aula 01 (síncrona e assíncrona): 24/08 – Compensação de conteúdo: Apresentação da disciplina • Aula 02 (síncrona e assíncrona): 26/08 - Compensação de conteúdo: Part-of-speech tagging • Aula 03 (síncrona e assíncrona): 31/08 - Idem • Aula 04 (síncrona): 02/09 – Normalização, tarefas de pré/pós-processamento; definição do Trabalho1 • Feriado: 07/09 • Aula 05 (síncrona): 09/09 - N-gram language models • Aula 06 (síncrona): 14/09 - Idem • Aula 07 (síncrona): 16/09 - Naive bayes models

UFRN/DIMAp

	<ul style="list-style-type: none">• Aula 08 (síncrona): 21/09 – Idem• Aula 09 (síncrona): 23/09 - Regressão logística• Aula 10 (síncrona): 28/09 - Idem• Aula 11 (síncrona): 30/09 - Hidden Markov models• Aula 12 (síncrona): 05/10 - Idem• Aula 13 (síncrona e assíncrona): 07/10 - Vector space models• Aula 14 (síncrona): 12/10 - Idem• Aula 15 (síncrona): 14/10 - Idem• Aula 16 (síncrona): 19/10 - Idem• Aula 17 (síncrona): 21/10 - Neural language models• Aula 18 (síncrona): 26/10 - Idem• Aula 19 (síncrona): 28/10 - Revisitando POS tagging• Feriado: 02/11• Aula 20 (síncrona): 04/11 - Idem• Aula 21 (síncrona): 09/11 - RNN• Aula 22 (síncrona): 11/11 - Idem• Aula 23 (síncrona e assíncrona): 16/11 - Execução do projeto final• Aula 24 (assíncrona): 18/11 - Idem• Aula 25 (assíncrona): 23/11 - Idem• Aula 26 (assíncrona): 25/11 - Idem• Aula 27 (assíncrona): 30/11 - Idem• Aula 28 (síncrona): 02/12 - Idem• Aula 29 (síncrona): 07/12 - Idem• Aula 30 (síncrona): 09/12 - Conclusão da disciplina• Aula 31 (síncrona): 14/12 - Prova de Reposição <p>O projeto final corresponderá a tópicos em parsing e aplicações de PLN mencionadas no final da ementa acima. Cada aluno ou dupla de alunos escolherá um tópico e desenvolverá o projeto sobre ele. Algumas das aulas correspondentes ao projeto serão ministradas no horário da grade para discutir os tópicos dos projetos, trocar ideias e resolver dúvidas, mas elas não são obrigatórias. Estas aulas vão depender da demanda. Se o aluno estiver presente na sala ele ganhará presença, caso contrário ele ou sua dupla poderão reunir-se separadamente com o professor de forma remota e isto corresponderá à presença. Quando não houver aula todas as reuniões serão pelos grupos. A outra maneira de auferir presenças nas aulas assíncronas relativas ao projeto final é apresentando por escrito um <i>progress report</i> sobre o projeto desde a aula anterior. A comunicação para esclarecer sobre cada momento será dinâmica pelos canais de comunicação usados na disciplina</p>
--	--

Datas e horários das atividades síncronas	Em caso de atividade síncrona, indicar o momento em que os encontros acontecerão, de acordo com o cronograma. Essa informação é importante para que os estudantes possam avaliar sua participação na turma.
	Vale salientar que qualquer atividade síncrona deve acontecer no horário previamente cadastrado para a turma, conforme indicado no Art. 3º, §2º "Para as atividades de interação online síncronas com os discentes, previstas nos planos de curso, os docentes deverão respeitar os dias e horários registrados para a turma no SIGAA".
	Veja calendário acima

UFRN/DIMAp

Compensação de Conteúdo	Descrição de como o conteúdo já ministrado nas semanas antes da suspensão das aulas será compensado, quando for o caso. <i>(Art. 3º §5º "Para turmas já iniciadas, deverá ser realizada a compensação de conteúdo.")</i>
	Inicialmente planeja-se três aulas de compensação de conteúdo, conforme calendário, mas isto pode variar dependendo dos alunos. Se forem apenas os mesmos alunos que já estavam matriculados, provavelmente será apenas uma semana. Se houver alunos novos, poderá ser mais de 3 aulas. A ideia também neste caso é apresentar a revisão em aula, de forma síncrona, mas também incentivar a que tenhamos sessão a seguir com os alunos novos individualmente ou em grupo para ver se conseguiram acompanhar.

Detalhamento dos recursos didáticos a serem utilizados	Descrição dos recursos didáticos a serem utilizados pelo docente para a realização das atividades.
	Para divulgação de notas e outras comunicações formais: sigaa Para entrega de trabalhos: salas de entrega do sigaa ou do moodle Para disponibilização de materiais diversos: moodle (Anteriormente já vinham sendo usados sigaa e moodle desta forma) Para aulas síncronas: Google Meet Para demonstrações/apresentações de trabalhos: Google Meet Para comunicação ágil com alunos: Discord

Recursos necessários para o Acompanhamento da turma pelo discente	Descrição dos recursos necessários para que o discente possa acompanhar as atividades da turma de forma adequada, por exemplo, plataformas de hardware e/ou software, requisitos computacionais e demais recursos necessários para realizar as atividades programadas.
	Computador, preferencialmente com Linux instalado, com acesso à Internet. Provavelmente precisará instalar Python e alguns outros softwares de acesso livre.

Materiais e Referências	Descrição dos materiais próprios ou de curadoria a serem utilizados para a realização das atividades, explicitando a forma de disponibilização para os discentes.
	Indicar referências a serem utilizadas para a realização das atividades, dando preferência a materiais gratuitos, online e acessíveis aos discentes.
	<i>(Art. 5º Os materiais didáticos deverão ser disponibilizados pelos docentes durante todo o período, considerando as limitações das condições de isolamento social impostas pela pandemia da COVID-19.</i> <i>§2º Quando necessário, os materiais utilizados nas atividades e/ou a forma de comunicação devem ser adaptados, de forma a atender discentes com algum tipo de deficiência e/ou com necessidades educacionais específicas, considerando as orientações da Secretaria de Inclusão e Acessibilidade (SIA).)</i>

	<p>BIBLIOGRAFIA BÁSICA:</p> <ol style="list-style-type: none">1. Jurafsky & Martin. Speech and Language Processing. 2 ed. Pearson. 2008. (Há um draft da 3ª ed em https://web.stanford.edu/~jurafsky/slp3/ed3book.pdf)2. Manning & Schütze. Foundations of Statistical Natural Language Processing. MIT Press. 1999.3. Sites de "recursos" de PLN
	<p>BIBLIOGRAFIA COMPLEMENTAR:</p> <ol style="list-style-type: none">1. Bender, Emily M. Linguistic Fundamentals for Natural Language Processing: 100 Essentials from Morphology and Syntax. Morgan & Claypool. 20132. Goldberg, Yoav. Neural Network Methods for Natural Language Processing. Synthesis Lectures on Human Language Technologies 10 (1), 1-309. 2017.3. Bird, Steven; Klein, Ewan; Loper, Edward. Disponível em https://www.nltk.org/book/4. NLTK group. NLTK – Natural Language Toolkit. Disponível em https://www.nltk.org/5. Stanford CoreNLP. Natural Language Software. Disponível em https://stanfordnlp.github.io/CoreNLP/
	<p>A principal referência é o livro de Jurafsky e Martin, cuja edição 3 está com draft disponível na Internet há vários anos, e é constantemente atualizada. Também são usados diversos materiais disponíveis pela Internet em particular corpora anotados e suas documentações, ferramentas de PLN, etc.</p>

Informações adicionais:	Acrescente aqui informações relevantes sobre o seu Plano de Curso e o desenvolvimento das atividades da turma.

Em, 01 de agosto de 2020.

Nome e assinatura do(a) docente