



UNIVERSIDADE FEDERAL DE OURO PRETO  
PRÓ-REITORIA DE GRADUAÇÃO  
PLANO DE ENSINO



|  |   |   |
|--|---|---|
| Nome do Componente Curricular em português:<br>Mineração Web   |   | Código:<br>BCC505                             |
| Nome do Componente Curricular em inglês:<br>Web Mining   |   |   |
| Nome e sigla do departamento:<br>Departamento de Computação (DECOM)  |   | Unidade acadêmica:<br>ICEB                    |
| Nome do docente:<br>Álvaro Rodrigues Pereira Júnior  |   |   |
| Carga horária semestral<br>60 horas  | Carga horária semanal teórica<br>2 horas/aula | Carga horária semanal prática<br>2 horas/aula |
| Data de aprovação na assembleia departamental:   |   |   |
| Ementa:<br>Classificação de texto; classificação em espaço vetorial; agrupamento; aprendizado de ranking; análise de links; sistemas de recomendação; análise de sentimento; coleta de dados Web; e aplicações de mineração na Web.  |   |   |
| Conteúdo programático:<br>- Introdução à mineração web e à ferramenta de apoio<br>- Coleta de dados web<br>- Extração de dados estruturados<br>- Fundamentos de aprendizado de máquina<br>- Fundamentos de recuperação de informação e busca web<br>- Fundamentos de mineração de dados<br>- Fundamentos de análise de redes complexas (redes sociais)<br>- Fundamentos de mineração de dados de uso e análise de sentimentos        |   |   |
| Objetivos:<br>Introduzir os conceitos da área de mineração web, apresentando um conjunto de técnicas usadas no processo de descoberta de conhecimento a partir de dados web, seja na etapa de coleta dos dados web, de extração e estruturação dos dados, até a etapa de mineração em si e logo de análise dos dados. Permitir que o estudante tenha acesso a um ferramental adequado para exploração prática dos conceitos da área. |   |   |
| Metodologia:<br>Considerando as características específicas que precisam ser consideradas para este curso em um formato especial e a distância, e atendendo ao pedido de Prograd em evitar atividades síncronas, este curso será oferecido integralmente em formato assíncrono, da seguinte forma:<br><br>Com relação ao conteúdo teórico, este será oferecido através de vídeo-aulas pelo professor,                                |   |   |

que também disponibilizará, para cada tema em estudo, uma série de artigos e referências digitais, em especial vídeo-aulas de outros professores, após curadoria. A cada semana uma nova aula teórica será apresentada.

Para o bom aprendizado do conteúdo teórico, o professor manterá um canal de comunicação simplificado com a turma, através de um grupo de Whatsapp, onde os alunos poderão lançar suas dúvidas, discutir com os colegas, e ter o retorno por parte do professor. Cada ponto de dúvida levantado será levado para um mural na página do curso no Moodle, para que os temas fiquem organizados e possam ser acessados por outros estudantes de forma estruturada.

A disciplina conta com 50% de sua carga-horária como sendo prática. Será utilizada uma plataforma de mineração de dados web para facilitar o processo de coleta e extração de dados, onde cada estudante deverá propor e construir uma aplicação de mineração web. Portanto, cada estudante fará o seu trabalho prático individualmente, ainda que possam haver trabalhos que façam parte de um mesmo projeto de aplicação.

A aplicação a ser desenvolvida por cada estudante deve aplicar com certa profundidade ao menos um conceito aprendido no curso sobre mineração web (ex. Sistemas de recomendação, análise de sentimentos, análise de redes complexas, regras de associação, entre outros).

Os estudantes reportarão os seus resultados parciais em cada semana do curso. Caberá ao professor auxiliar cada estudante no desenvolvimento do seu projeto, organizando o cronograma de entregas e, a cada semana, avaliar e responder ao estudante quanto à qualidade do desenvolvimento do seu projeto, com o objetivo de ter um trabalho final consistente.

Além do estudante reportar semanalmente o status do seu projeto de forma textual, este deverá apresentar duas *pitches* (apresentação que sumariza o que foi feito e tem menos de 5 minutos), disponibilizada para toda a turma em formato de vídeo digital. Este evento acontecerá no meio do semestre e ao final do semestre.

Sempre que o professor observar que um estudante não está evoluindo bem na sua atividade, deverá ser marcado um horário para uma reunião por vídeo-conferência (respeitando a janela de horário da disciplina) com este estudante, respeitando a sua disponibilidade de acesso a internet ou outros recursos que possa necessitar. Ainda que esta seja uma atividade síncrona, ela acontece apenas entre o professor e um dos estudantes. Portanto há flexibilidade para ajuste de horário que funcione bem para o estudante.

Atividades avaliativas:

Parte teórica: 30%

O estudante será avaliado a partir dos seguintes critérios:

- Demonstração de conhecimento a partir de sua interação com a classe através dos canais de comunicação (Moodle, whatsapp ou outros canais);
- Demonstração de conhecimento a partir do desenvolvimento de seu trabalho prático e de seu resultado final.

Parte prática: 70%

- Qualidade do projeto de trabalho prático;
- Nível de profundidade da aplicação dos conceitos de mineração web no trabalho prático;
- Qualidade das entregas parciais;
- Qualidade da entrega final;
- Resultado final: o trabalho prático tem aplicação real? É possível que o trabalho seja estendido futuramente para geração de impacto socioeconômico, visando resolver problemas reais?

#### Cronograma:

O conteúdo teórico semanal será o seguinte:

Semana 1: Introdução à mineração web e à ferramenta de apoio

Semana 2: Coleta de dados web

Semana 3: Extração de dados estruturados

Semana 4: Fundamentos de aprendizado de máquina

Semana 5: Fundamentos de recuperação de informação e busca web

Semana 6: Fundamentos de mineração de dados

Semana 7: Fundamentos de análise de redes complexas (redes sociais)

Semana 8: Fundamentos de mineração de dados de uso e análise de sentimentos

As atividades avaliativas acontecerão semanalmente, a partir das entregas parciais e da interação do estudante com a disciplina através de seus canais de comunicação, conforme apresentado acima.

O exame especial será uma nova tarefa ou melhoria do seu trabalho prático, que deverá ser realizada a partir do ponto em que estiver desenvolvido. Será realizado durante a semana subsequente à última semana de aula. O estudante deverá desenvolver algum tipo de melhoria no projeto, tendo em vista que para estar em exame especial, não foram alcançados os objetivos propostos no projeto apresentado por ele no início do curso.

#### Bibliografia básica:

LIU, Bing. Web data mining: exploring hyperlinks, contents, and usage data. 2<sup>nd</sup> Ed. New York: Springer, 2011. ISBN 978-3-642-19460-3

<https://www.passeidireto.com/arquivo/4007973/web-data-mining-bing-liu>

CHAKRABARTI, Soumen. Mining the Web: discovering knowledge from hypertext data. San Francisco, CA: Morgan Kaufmann Publishers, 2003.

WITTEN, Ian; FRANK, Eibe; Hall, Mark. Data Mining: Practical Machine Learning Tools and Techniques. 3rd Ed. Morgan Kaufmann Publishers, 2011.

<https://www.passeidireto.com/arquivo/16896806/data-mining-3rd-edition?q=data%20mining%20witten&tipo=1>

Bibliografia ou referências complementares disponíveis na web:

- Página web do Livro Web Data Mining. Bing Liu. Second Edition.  
<https://www.cs.uic.edu/~liub/WebMiningBook.html>
- Página pessoal do prof. Bing Liu. <https://www.cs.uic.edu/~liub/>
- Página web do Livro Mining the Web - Discovering Knowledge from Hypertext Data.  
<https://www.cse.iitb.ac.in/~soumen/mining-the-web/>
- Página pessoal do prof. Soumen Chakrabarti. <https://www.cse.iitb.ac.in/~soumen/>
- Web Data Mining. <https://www.youtube.com/playlist?list=PL0174E49C0E0DD5C8>
- Web Mining. <https://www.youtube.com/playlist?list=PL0174E49C0E0DD5C8>
- Introduction to Web Mining. [https://www.youtube.com/watch?v=dXQXJYce\\_rM](https://www.youtube.com/watch?v=dXQXJYce_rM)
- Basics of Data Mining. <https://www.youtube.com/watch?v=u2oSiVOQRmg>
- The State of Sentiment Analysis. Prof. Bing Liu.  
[https://www.youtube.com/watch?v=6ki\\_F\\_5a5vg](https://www.youtube.com/watch?v=6ki_F_5a5vg)