

Web Mining [WEM]

Responsable Raileanu Laura Elena

MRU TIC / HEIG-VD

Domaines de spécialisation CEN, TIN, TIC

Capacité d'accueil min. 5

Capacité d'accueil max. 40

Résumé Note: Ce cours sera enseigné par Mmes les professeures dr. Nastaran Fatemi et dr. Laura Elena Raileanu.

En février 2009, Tim Berners Lee annonçait que le Web n'était plus le "Web des documents", mais le "Web des données". En effet, le nouveau défi des applications orientées Web sera l'extraction des nouvelles connaissances à partir des teras octets de données distribuées de manière hétérogène. Ainsi, des outils performants devront être mis à disposition des utilisateurs du Web et des entreprises pour répondre de manière performante et intelligente à leurs besoins. Des initiatives majeures comme le "Semantic Web" et le "Linked Data", des applications récentes, comme par exemple "Wolfram Alpha", "OpenCalais" ou les "Rich snippets" de Google constituent des éloquentes exemples dans cette direction.

Le Web Mining représente l'ensemble des technologies qui permettent de découvrir des informations pertinentes et des nouvelles connaissances à partir des données du Web.

Nous commencerons le cours par présenter le domaine du Web Mining, et le situer par rapport aux autres domaines de recherches. Nous poserons ensuite la problématique du Web Mining, en la déclinant en trois catégories : le "Web content Mining", le "Web usage Mining" et le "Web structure Mining".

Le "Web content Mining" consiste dans l'application des techniques de data mining aux données pour extraire des informations structurées et ayant un contenu sémantique. Le "Web usage Mining" est devenu un outil efficace pour le domaine de business intelligence. L'application des techniques appropriées de data mining sur plusieurs sources de données et sur les actions des utilisateurs permet de comprendre et prédire le comportement des clients et de réaliser de la publicité ciblée, en suscitant un énorme intérêt commercial. En utilisant la théorie des graphes, le "Web structure Mining" permet d'analyser les liens entre différentes ressources et utilisateurs.

Pour chacune de ces catégories, nous présenterons la spécificité des données, des traitements et des techniques utilisées. Nous aborderons ensuite la complexité du web mining, du point de vue opératoire et systémique. Nous illustrerons chacune des parties présentées par des exemples d'applications. Nous terminerons le cours par les pistes de recherches actuellement explorées et leurs perspectives d'application.

Web Mining [WEM]

Contenu

| Sujet | Temps [%] |
|---|-----------|
| 1. Le Web mining, ses objectifs et son positionnement | 5 |
| - Un processus KDD sur des données Web | 5 |
| - Catégories du Web Mining et les domaines connexes | 5 |
| 2. Le Web structure Mining | 5 |
| - Les données utilisées (liens inter sites) | 5 |
| - Les pré-traitements | 5 |
| - Les méthodes (utilisation des graphes) | 5 |
| - Exemple d'application | 5 |
| 3. Le Web content Mining | 5 |
| - Les données utilisées (liens intra site et inter sites, balises de mise en forme, contenu multimédia) | 5 |
| - Les pré-traitements (structuration automatique des données non-structurées) | 5 |
| - Les méthodes (méthodes de l'apprentissage automatique) | 5 |
| - Exemple d'application | 5 |
| 4. Le Web usage Mining | 5 |
| - Les données utilisées (les cookies, les fichiers logs, les données d'enrichissement) | 5 |
| - Pré-traitements (structuration des données (fichiers logs)) | 5 |
| - Les méthodes (analyse exploratoire, règle d'associations, analyse de chemins) | 5 |
| - Exemple d'application | 5 |
| 5. Le Web mining face à la complexité opératoire et systémique | 5 |
| - Conclusions et perspectives | 5 |

Connaissances préalables

Bases de données
Connaissances d'un langage de programmation

Méthodes d'enseignement

| Mode | Périodes d'enseignement | Volume de travail (en heures) |
|--------------------------|-------------------------|-------------------------------|
| Exposés | 21 | 45 |
| Exercices | 0 | 0 |
| Travaux pratiques | 21 | 45 |
| TOTAL | 42 | 90 |
| Crédits ECTS | | 3 |

Évaluation

Examen écrit

Pondération de l'examen

50 %

Web Mining [WEM]

Compétences visées

| | |
|---|------------|
| Gérer le projet | 25% |
| Sait choisir et appliquer la méthode adéquate de gestion de projet, pour des projets de complexité moyenne | X |
| Sait identifier les contraintes économiques et les formuler (business plan) | X |
| Sait exploiter les ressources internes et identifier les ressources externes permettant de mettre en oeuvre une solution | |
| Est capable de s'intégrer dans un groupe; est en mesure d'animer, motiver et convaincre les membres du groupe | X |
| A le sens de l'initiative personnelle et des responsabilités | X |
| Analyser et spécifier des produits / services | 25% |
| Est capable d'analyser les besoins du client dans le domaine de spécialisation et sait traduire les exigences et contraintes dans le contexte technico-scientifico-économique et environnemental adéquat | X |
| Est capable de spécifier, planifier, concevoir et mettre en oeuvre des architectures de systèmes spécifiques au domaine de spécialisation, en intégrant des composants hétérogènes et en respectant les exigences d'interopérabilité et d'évolutivité des systèmes, ainsi que les normes et standards | X |
| Est capable de mener des études de faisabilité et de proposer des services de conseil | X |
| Est capable de superviser et analyser (monitoring) la sécurité d'un système IT et développer des tableaux de bord renseignant sur l'état du système | |
| Est capable d'effectuer une analyse du risque IT et sait choisir la méthode adéquate et, le cas échéant, l'adapter ou en développer une nouvelle | |
| Est capable de spécifier, dans un cahier des charges, les besoins du client, après les avoir traduits dans le contexte technico-économique adéquat | |
| Est en mesure de proposer et comparer des solutions et peut justifier un choix avec des arguments techniques, économiques, organisationnels ou environnementaux appropriés | X |
| Est capable de se mettre à la place de l'utilisateur pour concevoir un produit répondant à ses attentes | X |
| Développer et réaliser | 25% |
| Sait choisir et mettre en oeuvre efficacement un outil de modélisation dans son domaine de spécialisation | X |
| Est capable de choisir et mettre en oeuvre efficacement une approche d'aide à la décision pour résoudre des problèmes complexes et, le cas échéant, de l'adapter ou en développer une nouvelle | X |
| Est capable de choisir et mettre en oeuvre efficacement une méthode d'optimisation et, le cas échéant, de l'adapter ou en développer une nouvelle | X |
| Est capable de choisir et mettre en oeuvre efficacement une méthode de gestion et de configuration de réseaux et de services | |
| A appris à comparer entre elles diverses méthodes de recherche et de traitement de l'information multimédia et est capable d'en développer de nouvelles | X |
| A appris à comparer entre elles diverses méthodes de développement logiciel, de gestion de versions, de gestion de problèmes, de automatisée de logiciel et est capable de les appliquer, les adapter ou d'en développer de nouvelles | X |
| Sait appliquer les bonnes pratiques et modèles de conception (design patterns) pour des systèmes logiciels | |
| Sait utiliser à bon escient les concepts et techniques d'ingénierie et de stockage de l'information | X |
| Est capable d'évaluer et choisir des systèmes de transport (SAN, WAN, ?) et serveurs de stockage de l'information multimédia | |
| Est capable de proposer des approches innovantes pour la réalisation d'interfaces d'utilisateur adaptatives et adaptables en fonction des besoins et des profils des utilisateurs, en adoptant une approche ergonomique | |
| Est capable de proposer des approches innovantes pour la réalisation d'interfaces adaptatives en fonction du contexte (p. ex. drivers, type de réseau) | |
| Maîtrise les technologies de simulation graphiques tri-dimensionnelles et réalité virtuelle, p. ex. les GIS (Geographic Information Systems) | |
| Sait évaluer et choisir une méthode de traitement de l'information multimédia appropriée | |
| Connait les principes de l'informatique pervasive (ubiquitous computing) et sait les appliquer pour concevoir des solutions d'interaction homme-machine efficaces | |
| Connait les techniques de parallélisation logicielles et matérielles et de distribution des processus et des données | |
| Sait comparer les méthodes de co-design et est en mesure de choisir la méthode appropriée | |
| Est capable de mettre en oeuvre un outil de simulation de système complexe et d'optimiser son architecture, sa performance (p. ex. qualité de service) | |
| Est capable de concevoir, vérifier, réaliser et valider un système numérique | X |
| Sait appliquer les techniques de poly-publishing et de cross-média | |
| Est capable de développer, porter, adapter des composants logiciels de bas niveau (bootstrap, moniteur, driver, os, etc.) sur différentes architectures, en maîtrisant les aspects liés aux interactions logiciel-matériel | |
| Est capable de développer de nouvelles applications en respectant les contraintes propres aux environnements mobiles (os, transmission, consommation, interfaces, etc.) | |
| Est capable de modéliser un système physique en vue d'une implémentation informatique | |
| Sait appliquer des méthodologies de travail appropriées et organiser son temps | |
| A été sensibilisé aux règles d'éthique et du développement durable | X |
| Valider, améliorer et disséminer | 25% |
| Sait choisir et mettre en oeuvre efficacement un outil de test et de validation | |
| A appris à auditer un système d'information et est capable de proposer des mesures appropriées pour son amélioration | X |
| A appris à auditer la sécurité d'un système IT et est capable de proposer des mesures appropriées pour son amélioration | |
| A appris à auditer l'architecture d'un système de communication et est capable de proposer des mesures appropriées pour son amélioration | |
| A appris à auditer une architecture logicielle et le code y relatif et est capable de proposer des mesures appropriées pour son amélioration | |
| Est capable de choisir et mettre en oeuvre efficacement une approche d'ontologie informationnelle et de gestion de connaissances et, le cas échéant, de l'adapter ou en développer une nouvelle | |
| Est capable de concevoir et réaliser une plate-forme d'essai permettant de valider des architectures de systèmes ainsi que des composants matériels ou logiciels et d'optimiser leur fonctionnement | |
| Est en mesure d'assurer la veille technologique dans son domaine et d'intégrer les connaissances nouvelles | X |
| Sait rédiger, présenter, communiquer et convaincre de manière pertinente | X |
| Est intégré dans des réseaux professionnels lui facilitant les échanges d'information, les expériences et la veille technologique | |
| Est en mesure d'acquérir de façon autonome des connaissances et compétences nouvelles | X |