

POSTACADEMISCHE OPLEIDING

BIG DATA

14 maart 2019 – 27 juni 2019



UNIVERSITEIT
GENT

Deze opleiding is breed opgevat en behandelt de belangrijkste aspecten van datavergaring, -beheer, -analyse en -presentatie. Zo krijg je een totaalinzicht dat je beter in staat zal stellen om jouw Big Data efficiënt aan te wenden ten voordele van jouw onderneming. De aangeleerde technische inzichten worden aangevuld met een overzicht van de belangrijkste juridische en ethische aspecten, wat noodzakelijk is om Big Data op een maatschappelijk verantwoorde manier te gebruiken.

Big Data kunnen worden omschreven als gegevenscollecties die niet efficiënt met traditionele gegevensbeheer en -verwerkingstechnieken kunnen worden behandeld. Bepalende factoren daarbij zijn de grotere datavolumes, de grotere snelheden waarmee de data worden aangeboden en de grotere variëteit aan dataformaten en de kwaliteit van de data. De tendens naar Big Data wordt gevoed door de almaar groeiende beschikbaarheid van digitale informatie uit nieuwsbronnen, multimedia, sensors, ... en gaat gepaard met nieuwe uitdagingen om deze data efficiënt te kunnen verzamelen, opslaan, beheren, analyseren en presenteren.

Het inzetten van geavanceerde technologieën die specifiek zijn afgestemd op het verwerken van zeer grote hoeveelheden data, kan bedrijven helpen om beter tegemoet te komen aan de steeds groter wordende informatienoden die vaak vereist zijn om gegevensanalyse nog beter te kunnen onderbouwen. Een beter inzicht in de beschikbare data en een optimale exploitatie ervan levert de beste garantie om met meer kennis van zaken belangrijke beslissingen te onderbouwen en daar dan ook een concurrentieel voordeel mee te behalen.

DOELPUBLIEK

U krijgt inzicht in de problematiek die gepaard gaat met Big Data en in de beschikbare ICT-oplossingen die momenteel voorhanden zijn. Er wordt aangetoond hoe de aangereikte oplossingen werken, wat hun beperkingen en voordelen zijn en waar en wanneer ze het beste kunnen worden ingezet.

Voor de lessen wordt bewust gekozen voor een sterke academische aanpak waarbij de accenten liggen op het verwerven van kennis in de breedte zonder daarbij productgebonden te zijn. Deze lessen worden aangevuld met een aantal gastpresentaties waarbij aandacht wordt besteed aan praktische voorbeelden.

De opleiding is dusdanig opgevat dat deze toegankelijk is voor iedereen die ietwat vertrouwd is met informatica. Er wordt gewerkt rond hoorcolleges die handelen rond vier thema's: gegevensbeheer, gegevensanalyse, visualisatie en juridische en ethische aspecten. Deze vier thema's worden aangevuld met twee praktische getuigenissen met demo's.

GETUIGSCHRIFT

U ontvangt een getuigschrift, indien u deelneemt aan de volledige opleiding en slaagt voor het bijbehorende examen.

WETENSCHAPPELIJKE COÖRDINATIE

Prof. dr. Guy De Tré, Vakgroep Telecommunicatie en Informatieverwerking, Universiteit Gent

LESGEVERS

- Michael Brands, Dynactionize NV
- Antoon Bronselaer, Vakgroep Telecommunicatie en Informatieverwerking, Universiteit Gent
- Hans Constandt, Ontoforce
- Peter Craddock, DLA Piper
- Thomas Demeester, Vakgroep Informatietechnologie, Universiteit Gent
- Wesley De Neve, Vakgroep Elektronica en Informatiesystemen, Universiteit Gent
- Guy De Tré, Vakgroep Telecommunicatie en Informatieverwerking, Universiteit Gent
- Jan Fostier, Vakgroep Informatietechnologie, Universiteit Gent
- Frédéric Godin, Vakgroep Elektronica en Informatiesystemen, Universiteit Gent
- Peter Lambert, Vakgroep Elektronica en Informatiesystemen, Universiteit Gent
- Eva Lievens, Vakgroep Metajuridica, Privaat- en Ondernemingsrecht, Universiteit Gent
- Pieter Vandekerckhove, Add Perspective
- Dirk Van den Poel, Vakgroep Marketing, Universiteit Gent
- Baptist Vandersmissen, Vakgroep Elektronica en Informatiesystemen, Universiteit Gent
- Katrien Verbert, Departement Computerwetenschappen, KU Leuven
- Ruben Verborgh, Vakgroep Elektronica en Informatiesystemen, Universiteit Gent
- Bruno Volckaert, Vakgroep Informatietechnologie, Universiteit Gent

MEER INFO EN INSCHRIJVEN

WWW.UGAIN.UGENT.BE/BIGDATA

PROGRAMMA

1. GEGEVENSBEHEER

Inleiding en NoSQL

In de introductie wordt aandacht besteed aan de oorsprong van de term Big Data. Aspecten zoals de interpretatie, het belang, de problematiek en de kritiek op Big Data worden besproken.

Daarna komen de verschillende vormen en karakteristieken (Volume, Variety, Velocity en Veracity) van Big Data aan bod. Er wordt gekeken naar de tekortkomingen en beperkingen van traditionele databanksystemen en er wordt dieper ingegaan op mogelijke oplossingen. Vervolgens worden de belangrijkste NoSQL databankoplossingen ('Not only' SQL) gesitueerd en bestudeerd. Zowel key/value stores, documentdatabanken, column stores als graafdatabanken worden daarbij behandeld.

Datakwaliteit

In deze les wordt een overzicht gegeven van de verschillende technieken waarmee men datakwaliteit kan meten. Er wordt vervolgens uitgelegd hoe de meetresultaten van de verschillende technieken geïnterpreteerd moeten worden en hoe ze verder kunnen worden gebruikt in bijvoorbeeld rapportering en strategische analyses. Nadien worden ook enkele technieken uitgelegd voor de verbetering van kwaliteit. Alle methoden worden toegelicht aan de hand van cases uit de praktijk.

Information retrieval en document classificatie

Eerst wordt er een inleiding gegeven over 'Information Retrieval' (IR), de technologie die toelaat om grote digitale collecties van ongestructureerde teksten efficiënt te doorzoeken. De bedoeling is eerder om een overzicht te geven van de basisconcepten zoals indexering en retrieval modellen, dan om hier heel technisch op in te gaan. Nadien worden de mogelijkheden bekeken om in dergelijke collecties structuur aan te brengen door de documenten op specifieke manieren te gaan groeperen. Er worden enkele basisconcepten aangebracht rond machinaal leren in het algemeen, en rond classificatie en clustering. Vervolgens worden een paar belangrijke algoritmen besproken, en toegelicht via enkele applicaties.

Linked Data

Wat komt er na Big Data, als we gegevens niet meer in één plek bij elkaar kunnen brengen omwille van praktische, legale of andere redenen? In deze les bekijken we Linked Data, een andere manier om met gegevens om te gaan waarin data inherent verspreid zit over een netwerk. We behandelen technologieën uit het semantisch web met het oog op het machine-leesbaar maken van data en informatie. We bestuderen de noodzaak van semantiek om die data aan elkaar te koppelen. Daarnaast komen ook de principes van Open Data aan bod, met als doel het uniform beschikbaar stellen van data via het web.

2. GEGEVENSANALYSE

Het MapReduce-programmeermodel

MapReduce is een generiek raamwerk dat bijzonder geschikt is om op eenvoudige wijze analyses uit te voeren op enorme hoeveelheden data, gebruik makend van een parallel computersysteem. De basisconcepten rond 'Mappers' en 'Reducers' komen aan bod, alsook enkele veelgebruikte ontwerppatronen. Deze laatste worden gestaafd aan de hand van eenvoudig te begrijpen voorbeelden. De Hadoop-implementatie van MapReduce wordt besproken, alsook het verwante Hadoop Distributed File System (HDFS). De doelstelling is dat deelnemers na deze sessie in staat zijn Hadoop MapReduce toe te passen op eigen problemen.

Gedistribueerde gegevensverwerking

Voor de efficiënte verwerking van Big Data is men grotendeels naar volledig gedistribueerde vormen van gegevensverwerking overgestapt. Tijdens deze lesavond wordt u wegwijs gemaakt in enkele van de belangrijkste architecturen voor gedistribueerde gegevensverwerking (stream-gebaseerde dataverwerking, Lambda architectuur, Kappa architectuur, Microservices architectuur, Zeta architectuur, etc.). De opbouw en werking van deze architecturen worden besproken, hoe deze geheel of gedeeltelijk mappen op bestaande technologieën / implementaties (Apache Storm, Apache Samza, Apache Spark, Apache Kafka, etc.) en wat hun belangrijkste voor- en nadelen zijn. Dit geheel wordt aangevuld met voorbeelden van gedistribueerde architecturen die technologie-reuzen zoals LinkedIn, Netflix, etc. geadopteerd hebben om hen om te laten gaan met de enorme hoeveelheid data die ze dagelijks moeten verwerken.

Deep learning

Kunstmatige neurale netwerken zijn in staat om het menselijk leerproces na te bootsen door het veranderen van de sterkte van gesimuleerde neurale verbindingen, een eigenschap die ervoor zorgt dat deze netwerken uiterst effectief zijn in het automatisch terugvinden van patronen in grote hoeveelheden data (deep learning). Dit heeft onlangs geleid tot een aantal doorbraken op het vlak van taalverwerking en audiovisuele analyse. Voortbouwend op een aantal basisconcepten uit het domein van machinaal leren, wordt er in deze les bijzondere aandacht besteed aan het gebruik van meerlagige neurale netwerken, alsook aan de technieken die de inzet van deze netwerkarchitecturen praktisch haalbaar hebben gemaakt. Vervolgens wordt er stilgestaan bij verschillende toepassingen op het vlak van taalverwerking en audiovisuele analyse, illustrerend hoe meerlagige neurale netwerken kunnen aangewend worden om kennis te extraheren uit grote hoeveelheden ruizige data. Tot slot wordt er eveneens een overzicht gegeven van toekomstige uitdagingen op het vlak van onderzoek en ontwikkeling in het domein van deep learning.

Big Data Analytics

- Inleiding tot Analytics (descriptive, predictive en prescriptive analytics);
- De Spark (Streaming) software stack (Berkeley Data Analytics Stack (BDAS));
- Gebruik van ML Pipelines (machine learning library) binnen Spark voor Analytics;
- Bespreking van enkele gebruikscases van de Spark software stack voor Analytics;
- Hands-on demo van een UGent-implementatie van een gevallenstudie.

3. VISUALISATIE

Getuigenissen uit de praktijk

De eerste spreker is Hans Constandt, CEO en medeoprichter van Ontoforce. Ontoforce is een Belgisch bedrijf dat oplossingen aanbiedt voor 'information flow' en management. In deze lezing wordt een gebruiksvriendelijk data zoekplatform voorgesteld dat gedreven is door semantische technologie en in staat is om intelligente links te bouwen tussen de steeds groeiende hoeveelheid van interne en externe gegevens. Tevens wordt een toepassing voor de farmaceutische industrie toegelicht.

De tweede spreker is Michael Brands, oprichter en CEO van Dynactionize, een Belgisch bedrijf dat een semantisch intelligent data platform ontwikkeld heeft dat naadloos gestructureerde, semi-gestructureerde en ongestructureerde data integreert. Het platform detecteert automatisch relevante directe en indirecte links tussen data elementen. De werking van die semantische intelligentie wordt geïllustreerd aan de hand van een aantal praktijktoepassingen uit de politiewereld, de gezondheidszorg, de logistiek en uit customer intelligence.

Visualisatie

Een uitgelezen manier om mensen te helpen om Big Data te exploreren en te begrijpen, is het visualiseren van de data: we zijn immers vaak erg goed in staat om patronen, tendensen, uitschieters, ... te begrijpen met behulp van visualisaties. Mede aan de hand van een groot aantal concrete voorbeelden wordt uiteengezet hoe een goede interactieve visualisatie kan worden opgebouwd, wat de typische misvattingen zijn, hoe visualisaties kunnen misbruikt worden, ... Er wordt ook een overzicht gepresenteerd van een aantal typische technieken en hulpmiddelen voor interactieve informatie visualisatie.

4. JURIDISCHE EN ETHISCHE ASPECTEN

Juridische aspecten

U wordt wegwijs gemaakt in de juridische uitdagingen rond Big Data projecten zoals privacy en gegevensbescherming, consumentenbescherming, intellectuele eigendomsrechten, discriminatie en andere relevante topics. U krijgt tevens een aantal vuistregels omtrent het op een juridisch correcte manier aanvatten en uitvoeren van Big Data projecten. Aan de hand van use cases worden een paar veel voorkomende juridische problemen uit de praktijk en hun mogelijke oplossingen nader toegelicht.

Ethische aspecten

Big Data is een term die verwijst naar technologische processen die vandaag deel zijn van onze meest intieme sfeer. Bestaat er een ethiek van Big Data? Hoe kunnen we Big Data op een ethisch verantwoorde manier gebruiken? Tijdens deze lesavond staan de volgende leerdoelen voorop:

- Begrijpen wat ethiek van een technologie is;
- Begrijpen dat Big Data niet ethisch neutraal is en hoe ontwikkelaars bepaalde waarden mee "programmeren";
- Verschillende ethische perspectieven kunnen toepassen op Big Data technologie.

PRAKTISCH

Prijs

Deelnameprijs omvat lesgeld, hand-outs, frisdranken, koffie en broodjes.

Betaling geschiedt na ontvangst van de factuur.

Alle facturen zijn betaalbaar dertig dagen na dagtekening.

Alle vermelde bedragen zijn vrij van BTW.

Module 1 GEGEVENSBEHEER	€ 600,-
Module 2 GEGEVENSANALYSE	€ 600,-
Module 3 VISUALISATIE	€ 300,-
Module 4 JURIDISCHE EN ETHISCHE ASPECTEN	€ 450,-
Volledige opleiding	€ 1.755,-

Voor iedere module kan er afzonderlijk ingeschreven worden.

Korting

- Indien minstens één deelnemer van een bedrijf inschrijft voor de volledige opleiding wordt voor alle bijkomende gelijktijdige inschrijvingen van hetzelfde bedrijf een korting van 20% verleend. Facturatie geschiedt dan d.m.v. een gezamenlijke factuur.
- Aangepaste prijzen voor personeel van UGent en geassocieerde hogescholen.
- Kortingen zijn niet cumuleerbaar.

Annulering

Raadpleeg onze annulatievoorwaarden op
www.ugain.ugent.be/annulatievoorwaarden

KMO-portefeuille

Universiteit Gent aanvaardt betalingen via de KMO-portefeuille (www.kmo-portefeuille.be; gebruik autorisatiecode DV.0103194).

Tijdstip en locatie

- De lessen worden gegeven van 18u tot 21u30, in 2 delen, gescheiden door een broodjesmaaltijd en vinden plaats aan de Universiteit Gent, UGent Academie voor Ingenieurs, Technologiepark 904 (nieuw nummer **60** vanaf januari 2019), 9052 Zwijnaarde.
- Data onder voorbehoud van wijzigingen om onvoorziene omstandigheden.

1. GEGEVENSBEHEER

14 maart 2019	Inleiding en NoSQL Guy De Tré
21 maart 2019	Datakwaliteit Antoon Bronselaer
28 maart 2019	Information retrieval en document classificatie Thomas Demeester
4 april 2019	Linked Data Peter Lambert en Ruben Verborgh

2. GEGEVENSANALYSE

25 april 2019	Het MapReduce-programmeermodel Jan Fostier
2 mei 2019	Gedistribueerde gegevensverwerking Bruno Volckaert
9 mei 2019	Deep learning Wesley De Neve, Frédéric Godin en Baptist Vandersmissen
16 mei 2019	Big Data Analytics Dirk Van den Poel

3. VISUALISATIE

23 mei 2019	Getuigenissen uit de praktijk Hans Constandt en Michael Brands
6 juni 2019	Visualisatie Katrien Verbert

4. JURIDISCHE EN ETHISCHE ASPECTEN

13 en 20 juni 2019	Juridische aspecten Peter Craddock, Eva Lievens en Frederik Ringoot
27 juni 2019	Ethische aspecten Pieter Vandekerckhove

Organisatie

Universiteit Gent

UGain (UGent Academie voor Ingenieurs)

Technologiepark 904 (nieuw nummer **60** vanaf januari 2019)

9052 Zwijnaarde

09 264 55 82

ugain@ugent.be - www.ugain.ugent.be

MEER INFO EN INSCHRIJVEN

WWW.UGAIN.UGENT.BE/BIGDATA



UNIVERSITEIT
GENT

FACULTEIT INGENIEURSWETENSCHAPPEN
EN ARCHITECTUUR

FACULTEIT
BIO-INGENIEURSWETENSCHAPPEN