

## Datakwaliteit (E018700)

**Cursusomvang** *(nominale waarden; effectieve waarden kunnen verschillen per opleiding)*

**Studiepunten 3.0** **Studietijd 90 u**

**Aanbodsessies in academiejaar 2025-2026**

A (semester 1) Engels Gent

**Lesgevers in academiejaar 2025-2026**

Bronselaer, Antoon TW07 Verantwoordelijk lesgever

**Aangeboden in onderstaande opleidingen in 2025-2026**

	stptn	aanbodsessie
<a href="#">Brugprogramma Master of Science in Bioinformatics (afstudeerrichting Engineering)</a>	3	A
<a href="#">Master of Science in Business Engineering (afstudeerrichting Data Analytics)</a>	3	A
<a href="#">Master of Science in Bioinformatics (afstudeerrichting Engineering)</a>	3	A
<a href="#">Master of Science in Business Engineering (afstudeerrichting Operations Management)</a>	3	A
<a href="#">Master of Science in Computer Science Engineering</a>	3	A
<a href="#">Master of Science in de industriële wetenschappen: informatica</a>	3	A
<a href="#">Master of Science in de informatica</a>	3	A
<a href="#">Uitwisselingsprogramma industriële wetenschappen: informatica</a>	3	A
<a href="#">Uitwisselingsprogramma informatica (niveau master)</a>	3	A

### Onderwijstalen

Engels

### Trefwoorden

Meten van datakwaliteit, consistentie, detectie en herstel van fouten, outlier detection

### Situering

Deze cursus is een specialisatiecursus waarin mechanismen voor het waarborgen van datakwaliteit centraal staan. Een eerste deel van de cursus behandelt verschillende methoden voor het meten van kwaliteit. In een tweede deel komen algoritmen aan bod die toelaten om fouten te detecteren, alsook algoritmen om deze fouten systematisch te gaan corrigeren. Een derde deel behandelt ontduubeling van data. Een vierde deel behandelt het detecteren van outliers en tot slot bekijken we in een vijfde en laatste deel een aantal specifieke kwaliteitsproblemen bij temporele data.

### Inhoud

- Inleiding tot datakwaliteit
- Meten van datakwaliteit: ordinale systemen, onzekerheidsmodellen en kost-gebaseerd meten
- Basis van constraint-gebaseerde formalismen
- Controle getallen
- Edit regels: foutlocalisatie, Fellegi-Holt model, FCF-algoritme
- Chase algoritme voor functionele afhankelijkheden
- Data ontduubeling: Fellegi-Sunter model, string vergelijking, samenvoegen van dubbele data
- Outlier detectie: afstandsgebaseerde modellen, pivot index, ruimtelijke partitionering, isolation forests
- Datakwaliteit in temporele data: trend decompositie, change detection, currency

### Begincompetenties

Vertrouwd zijn met gegevensstructuren en relationele databanken. Basiskennis van (Goedgekeurd)

programmeren.

### **Eindcompetenties**

- 1 Inzicht hebben in de basistechnieken voor het meten van datakwaliteit.
- 2 Begrijpen hoe consistentie kan worden afgedwongen en in staat zijn om dit toe te passen.
- 3 Begrijpen hoe een dataset kan worden ontdekt.
- 4 Begrijpen hoe outliers gezocht kunnen worden.
- 5 Inzicht hebben in de specifieke kwaliteitsproblemen bij temporele data.

### **Creditcontractvoorwaarde**

Toelating tot dit opleidingsonderdeel via creditcontract is mogelijk na gunstige beoordeling van de competenties

### **Examencontractvoorwaarde**

Dit opleidingsonderdeel kan niet via examencontract gevolgd worden

### **Didactische werkvormen**

Werkcollege, Hoorcollege

### **Studiemateriaal**

Type: Slides

Naam: Slides

Richtprijs: € 25

Optioneel: nee

Taal : Engels

Aantal slides : 440

Oudst bruikbare editie : Geen, slides worden jaarlijks up-to-date gehouden

Beschikbaar op Ufora : Ja

Online beschikbaar : Nee

Beschikbaar in de bibliotheek : Nee

Beschikbaar via studentenvereniging : Ja

### **Referenties**

- Ton De Waal, Jeroen Pannekoek en Sander Scholtus (2011). Handbook of Statistical Data Editing and Imputation, Wiley
- Wenfei Fan en Floris Geerts (2012). Foundations of Data Quality Management. Morgan & Claypool Publishers.

### **Vakinhoudelijke studiebegeleiding**

Alle oefeningenlessen worden begeleid door assistenten.

### **Evaluatiemomenten**

periodegebonden evaluatie

### **Evaluatievormen bij periodegebonden evaluatie in de eerste examenperiode**

Schriftelijke evaluatie

### **Evaluatievormen bij periodegebonden evaluatie in de tweede examenperiode**

Schriftelijke evaluatie

### **Evaluatievormen bij niet-periodegebonden evaluatie**

### **Tweede examenkans in geval van niet-periodegebonden evaluatie**

Niet van toepassing

### **Toelichtingen bij de evaluatievormen**

Periodegebonden evaluatie: schriftelijk examen (gesloten boek) bestaande uit oefeningen en open vragen die naar inzicht in de cursus peilen.

### **Eindscoreberekening**

100% schriftelijk examen