

Guia docent

2709952 - IQL - Introducció a la Lingüística Quantitativa

Última modificació: 30/01/2026

Unitat responsable: Facultat d'Informàtica de Barcelona
Unitat que imparteix: 723 - CS - Departament de Ciències de la Computació.

Titulació: MÀSTER UNIVERSITARI EN ENGINYERIA INFORMÀTICA (Pla 2012). (Assignatura optativa).
MÀSTER UNIVERSITARI EN INNOVACIÓ I RECERCA EN INFORMÀTICA (Pla 2012). (Assignatura optativa).
MÀSTER UNIVERSITARI EN CIÈNCIA DE DADES (Pla 2021). (Assignatura optativa).

Curs: 2025 **Crèdits ECTS:** 6.0 **Idiomes:** Anglès

PROFESSORAT

Professorat responsable:

Altres:

CAPACITATS PRÈVIES

Programació
Probabilitat i estadística bàsiques

COMPETÈNCIES DE LA TITULACIÓ A LES QUALS CONTRIBUEIX L'ASSIGNATURA

Transversals:

CTR3. TREBALL EN EQUIP: Ser capaç de treballar com a membre d'un equip, ja sigui com a un membre més, ja sigui realitzant tasques de direcció, amb la finalitat de contribuir a desenvolupar projectes d'una manera pragmàtica i amb sentit de la responsabilitat; assumir compromisos tenint en compte els recursos disponibles.

CTR4. ÚS SOLVENT DELS RECURSOS D'INFORMACIÓ: Gestionar l'adquisició, l'estructuració, l'anàlisi i la visualització de dades i d'informació de l'àmbit de l'enginyeria informàtica, i valorar de forma crítica els resultats d'aquesta gestió.

CTR6. RAONAMENT: Capacitat de raonament crític, lògic i matemàtic. Capacitat de resoldre problemes en la seva àrea d'estudi. Capacitat d'abstracció: capacitat de crear i utilitzar models que reflecteixin situacions reals. Capacitat de dissenyar i realitzar experiments senzills, i analitzar-ne i interpretar-ne els resultats. Capacitat d'anàlisi, de síntesi i d'avaluació.

CTR9. TERCERA LLENGUA: Conèixer una tercera llengua, preferentment l'anglès, amb un nivell adequat oral i escrit i en consonància amb les necessitats que tindran els titulats i titulades.

METODOLOGIES DOCENTS

Les sessions de teoria les durà a terme principalment el professor, ja sigui mitjançant la pissarra o projectant diapositives.

El treball de laboratori es farà davant de l'ordinador. S'espera que els estudiants treballin en els seus deures i el professor explicarà tot el necessari per seguir la classe al principi de la sessió. Cada sessió de laboratori anirà acompanyada d'una guia completa que descriurà el treball que cal fer.

El projecte de recerca es durà a terme sota la supervisió del professor.

Tot el material rellevant per al curs estarà disponible al Racó o al lloc web del curs.

OBJECTIUS D'APRENTATGE DE L'ASSIGNATURA

1. Conèixer els fonaments de la ciència i del mètode científic. Entendre la diferència entre hipòtesi i teoria, entre modelar i comprendre, entre descriure i explicar, entre manifestació i principi. Entendre el valor de la predicció i els tipus de predicció.
2. Conèixer les lleis estadístiques del llenguatge i entendre els seus orígens.
3. Conèixer i entendre els principis d'organització de les llengües i altres sistemes de comunicació
4. Conèixer els fonaments matemàtics de la lingüística quantitativa. Conèixer teoria de probabilitat bàsica i teoria de la informació.
5. Conèixer els mètodes d'anàlisi estadística de la lingüística quantitativa.
6. Aprendre a redactar un article científic. Saber distingir entre un informe de laboratori i un article de recerca. Aprendre a redactar un article científic en equip.

HORES TOTALES DE DEDICACIÓ DE L'ESTUDIANTAT

Tipus	Hores	Percentatge
Hores aprenentatge autònom	102,0	68.00
Hores grup gran	24,0	16.00
Hores grup petit	24,0	16.00

Dedicació total: 150 h

CONTINGUTS

Introducció a la Lingüística Quantitativa

Descripció:

Què és la lingüística quantitativa? Panoràmica de lleis lingüístiques, conceptes clau i problemes de recerca en lingüística quantitativa.

Llei de brevetat i problema de la compressió

Descripció:

La llei de brevetat en humans i altres espècies. Mètodes d'anàlisi de la llei de brevetat. Introducció a la teoria de la informació. Prediccions de la codificació òptima.

Teoria de la informació

Descripció:

Teoria de la informació clàssica i extensions per a sistemes de comunicació naturals.

Teoria de lleis potencials

Descripció:

Relacions entre lleis potencials. Deducció de lleis potencials. Mètodes d'anàlisi de lleis potencials.

Models de la llei de Zipf per a la freqüència de les paraules

Descripció:

Fites de DeBowski. Models clàssics. Models zipfians d'optimització de la comunicació.

L'estructura estadística de les seqüències simbòliques

Descripció:

Retorn de paraules. Correlacions en seqüències simbòliques. Persistència i antipersistència. Models basats en n-grames. Models generatius.

Sintaxi de dependències

Descripció:

Introducció a la sintaxi de dependències. Restriccions formals sobre estructures de dependències sintàctiques

Teoria d'ordre de les paraules

Descripció:

Principis d'ordre de les paraules. Prediccions. Ordre del subjecte (S), complement directe (O) i verb (V).

Construcció de teoria

Descripció:

El mètode científic. Una teoria general. Tancament

ACTIVITATS

Introducció

Descripció:

Introducció a la lingüística quantitativa. Introducció a l'assignatura

Objectius específics:

1, 2

Competències relacionades:

CTR4. ÚS SOLVENT DELS RECURSOS D'INFORMACIÓ: Gestionar l'adquisició, l'estructuració, l'anàlisi i la visualització de dades i d'informació de l'àmbit de l'enginyeria informàtica, i valorar de forma crítica els resultats d'aquesta gestió.

CTR6. RAONAMENT: Capacitat de raonament crític, lògic i matemàtic. Capacitat de resoldre problemes en la seva àrea d'estudi.

Capacitat d'abstracció: capacitat de crear i utilitzar models que reflecteixin situacions reals. Capacitat de dissenyar i realitzar experiments senzills, i analitzar-ne i interpretar-ne els resultats. Capacitat d'anàlisi, de síntesi i d'avaluació.

Dedicació: 8h 18m

Grup gran/Teoria: 3h

Aprenentatge autònom: 5h 18m

Llei de brevetat i el problema de la compressió

Objectius específics:

4, 5

Competències relacionades:

CTR4. ÚS SOLVENT DELS RECURSOS D'INFORMACIÓ: Gestionar l'adquisició, l'estructuració, l'anàlisi i la visualització de dades i d'informació de l'àmbit de l'enginyeria informàtica, i valorar de forma crítica els resultats d'aquesta gestió.

CTR6. RAONAMENT: Capacitat de raonament crític, lògic i matemàtic. Capacitat de resoldre problemes en la seva àrea d'estudi. Capacitat d'abstracció: capacitat de crear i utilitzar models que reflecteixin situacions reals. Capacitat de dissenyar i realitzar experiments senzills, i analitzar-ne i interpretar-ne els resultats. Capacitat d'anàlisi, de síntesi i d'avaluació.

Dedicació: 15h

Grup gran/Teoria: 3h 48m

Grup mitjà/Pràctiques: 1h

Grup petit/Laboratori: 3h 30m

Aprenentatge autònom: 6h 42m

Teoría de la informació

Objectius específics:

3, 4

Competències relacionades:

CTR6. RAONAMENT: Capacitat de raonament crític, lògic i matemàtic. Capacitat de resoldre problemes en la seva àrea d'estudi. Capacitat d'abstracció: capacitat de crear i utilitzar models que reflecteixin situacions reals. Capacitat de dissenyar i realitzar experiments senzills, i analitzar-ne i interpretar-ne els resultats. Capacitat d'anàlisi, de síntesi i d'avaluació.

Dedicació: 12h 06m

Grup gran/Teoria: 4h

Grup mitjà/Pràctiques: 1h

Aprenentatge autònom: 7h 06m

Teoria de lleis potencials

Objectius específics:

2, 4

Competències relacionades:

CTR4. ÚS SOLVENT DELS RECURSOS D'INFORMACIÓ: Gestionar l'adquisició, l'estructuració, l'anàlisi i la visualització de dades i d'informació de l'àmbit de l'enginyeria informàtica, i valorar de forma crítica els resultats d'aquesta gestió.

CTR6. RAONAMENT: Capacitat de raonament crític, lògic i matemàtic. Capacitat de resoldre problemes en la seva àrea d'estudi. Capacitat d'abstracció: capacitat de crear i utilitzar models que reflecteixin situacions reals. Capacitat de dissenyar i realitzar experiments senzills, i analitzar-ne i interpretar-ne els resultats. Capacitat d'anàlisi, de síntesi i d'avaluació.

Dedicació: 15h 36m

Grup gran/Teoria: 4h

Grup mitjà/Pràctiques: 1h

Grup petit/Laboratori: 3h 30m

Aprenentatge autònom: 7h 06m

Models de la llei de Zipf per a la freqüència de les paraules

Objectius específics:

1, 2, 3, 4

Competències relacionades:

CTR4. ÚS SOLVENT DELS RECURSOS D'INFORMACIÓ: Gestionar l'adquisició, l'estructuració, l'anàlisi i la visualització de dades i d'informació de l'àmbit de l'enginyeria informàtica, i valorar de forma crítica els resultats d'aquesta gestió.

CTR6. RAONAMENT: Capacitat de raonament crític, lògic i matemàtic. Capacitat de resoldre problemes en la seva àrea d'estudi. Capacitat d'abstracció: capacitat de crear i utilitzar models que reflecteixin situacions reals. Capacitat de dissenyar i realitzar experiments senzills, i analitzar-ne i interpretar-ne els resultats. Capacitat d'anàlisi, de síntesi i d'avaluació.

Dedicació: 11h 54m

Grup gran/Teoria: 4h

Grup mitjà/Pràctiques: 0h 48m

Aprenentatge autònom: 7h 06m

L'estructura estadística de les seqüències simbòliques

Objectius específics:

2, 3, 4, 5

Competències relacionades:

CTR4. ÚS SOLVENT DELS RECURSOS D'INFORMACIÓ: Gestionar l'adquisició, l'estructuració, l'anàlisi i la visualització de dades i d'informació de l'àmbit de l'enginyeria informàtica, i valorar de forma crítica els resultats d'aquesta gestió.

CTR6. RAONAMENT: Capacitat de raonament crític, lògic i matemàtic. Capacitat de resoldre problemes en la seva àrea d'estudi. Capacitat d'abstracció: capacitat de crear i utilitzar models que reflecteixin situacions reals. Capacitat de dissenyar i realitzar experiments senzills, i analitzar-ne i interpretar-ne els resultats. Capacitat d'anàlisi, de síntesi i d'avaluació.

Dedicació: 15h 06m

Grup gran/Teoria: 4h

Grup mitjà/Pràctiques: 1h

Grup petit/Laboratori: 3h

Aprenentatge autònom: 7h 06m

Sintaxi de dependències

Objectius específics:

2, 5

Competències relacionades:

CTR4. ÚS SOLVENT DELS RECURSOS D'INFORMACIÓ: Gestionar l'adquisició, l'estructuració, l'anàlisi i la visualització de dades i d'informació de l'àmbit de l'enginyeria informàtica, i valorar de forma crítica els resultats d'aquesta gestió.

CTR6. RAONAMENT: Capacitat de raonament crític, lògic i matemàtic. Capacitat de resoldre problemes en la seva àrea d'estudi. Capacitat d'abstracció: capacitat de crear i utilitzar models que reflecteixin situacions reals. Capacitat de dissenyar i realitzar experiments senzills, i analitzar-ne i interpretar-ne els resultats. Capacitat d'anàlisi, de síntesi i d'avaluació.

Dedicació: 12h 06m

Grup gran/Teoria: 4h

Grup mitjà/Pràctiques: 1h

Aprenentatge autònom: 7h 06m

Teoria d'ordre de les paraules

Objectius específics:

3, 4, 5

Competències relacionades:

CTR4. ÚS SOLVENT DELS RECURSOS D'INFORMACIÓ: Gestionar l'adquisició, l'estructuració, l'anàlisi i la visualització de dades i d'informació de l'àmbit de l'enginyeria informàtica, i valorar de forma crítica els resultats d'aquesta gestió.

CTR6. RAONAMENT: Capacitat de raonament crític, lògic i matemàtic. Capacitat de resoldre problemes en la seva àrea d'estudi. Capacitat d'abstracció: capacitat de crear i utilitzar models que reflecteixin situacions reals. Capacitat de dissenyar i realitzar experiments senzills, i analitzar-ne i interpretar-ne els resultats. Capacitat d'anàlisi, de síntesi i d'avaluació.

Dedicació: 15h 36m

Grup gran/Teoria: 4h

Grup mitjà/Pràctiques: 1h

Grup petit/Laboratori: 3h 30m

Aprenentatge autònom: 7h 06m

Construcció de teoria

Objectius específics:

1, 3

Competències relacionades:

CTR6. RAONAMENT: Capacitat de raonament crític, lògic i matemàtic. Capacitat de resoldre problemes en la seva àrea d'estudi. Capacitat d'abstracció: capacitat de crear i utilitzar models que reflecteixin situacions reals. Capacitat de dissenyar i realitzar experiments senzills, i analitzar-ne i interpretar-ne els resultats. Capacitat d'anàlisi, de síntesi i d'avaluació.

Dedicació: 8h 18m

Grup gran/Teoria: 3h

Aprenentatge autònom: 5h 18m

Projecte de recerca

Objectius específics:

1, 2, 3, 4, 5, 6

Competències relacionades:

CTR3. TREBALL EN EQUIP: Ser capaç de treballar com a membre d'un equip, ja sigui com a un membre més, ja sigui realitzant tasques de direcció, amb la finalitat de contribuir a desenvolupar projectes d'una manera pragmàtica i amb sentit de la responsabilitat; assumir compromisos tenint en compte els recursos disponibles.

CTR4. ÚS SOLVENT DELS RECURSOS D'INFORMACIÓ: Gestionar l'adquisició, l'estructuració, l'anàlisi i la visualització de dades i d'informació de l'àmbit de l'enginyeria informàtica, i valorar de forma crítica els resultats d'aquesta gestió.

CTR6. RAONAMENT: Capacitat de raonament crític, lògic i matemàtic. Capacitat de resoldre problemes en la seva àrea d'estudi. Capacitat d'abstracció: capacitat de crear i utilitzar models que reflecteixin situacions reals. Capacitat de dissenyar i realitzar experiments senzills, i analitzar-ne i interpretar-ne els resultats. Capacitat d'anàlisi, de síntesi i d'avaluació.

Dedicació: 36h

Aprenentatge autònom: 36h

SISTEMA DE QUALIFICACIÓ

La qualificació es fa mitjançant exàmens i informes sobre diverses tasques (pràctiques de laboratori i un projecte de recerca) al llarg del curs.

Hi haurà dos exàmens parcials que compten un 30% de la nota. S'espera que els estudiants lliurin 4 informes de treballs de laboratori aproximadament dues setmanes després de la sessió de laboratori corresponent, que compten fins al 30% de la nota final. Finalment, els estudiants hauran de lliurar un projecte de recerca al final del curs que representa el 40% de la nota final. El projecte de recerca és l'activitat més important i s'ha d'entendre com un projecte del curs (no com una pràctica més). Cal entendre les pràctiques com a un entrenament per al projecte de recerca.

Per tant, la fórmula per calcular la nota final és:

$$0,4 * (P1 + P2) + 0,2 * (L1 + L2 + L3 + L4) + 0,4 * CP$$

on P1 és la nota del primer examen parcial, P2 és la nota del segon examen parcial, Li correspon a la nota de la i-èssima pràctica de laboratori i RP és la nota del projecte de recerca.

BIBLIOGRAFIA

Bàsica:

- "Linguistic laws in biology". Linguistic laws in biology.
- Zipf, George K.. Human behavior and the principle of least effort.. Addison-Wesley Press, 1949.
- Best, Karl-Heinz; Rottmann, Otto. Quantitative linguistics, an invitation. Ram-Verlag, 2017. ISBN 9783942303514.
- Bentz, Christian. Adaptive languages: an information-theoretic account of linguistic diversity. 1st ed. 2018. ISBN 9783110557770.
- Tanaka-Ishii, Kumiko. Statistical universals of language: mathematical chance vs human choice. Springer, 2021. ISBN 9783030593773.
- Levshina, Natalia. Communicative efficiency: Language structure and use. Cambridge University Press, 2023.
- Baayen, R. Harald. Analyzing linguistic data: a practical introduction to statistics using R. Cambridge University, 2008. ISBN 9780521709187.
- Bunge, Mario. La Ciencia: su método y su filosofía. Ed. corr. y aum. Sudamericana, 1995. ISBN 9500710439.

Complementària:

- Herdan, Gustav. The calculus of linguistic observations. Reprint 2021. 1962. ISBN 9783112415443.
- Debowski, Lukasz. Information theory meets power laws: Stochastic processes and language models. Wiley, 2021.

RECURSOS

Enllaç web:

- <https://cqlab.upc.edu/biblio/laws/>. Laws of language outside human language. Statistical laws of language in the behavior of other species, genomes and beyond