



Πανεπιστήμιο Πειραιώς, Τμήμα Πληροφορικής
ΠΜΣ «Κυβερνοασφάλεια και Επιστήμη Δεδομένων»
ακαδ. έτος 2023-2024

ΔΙΑΧΕΙΡΙΣΗ ΜΕΓΑΛΩΝ ΔΕΔΟΜΕΝΩΝ

Διδάσκοντες: Νίκος Πελέκης, Γιώργος Παπαστεφανάτος
Εργαστηριακοί βοηθοί: Γ. Αλεξίου, Σ. Μαρούλης

Βιβλιογραφική εργασία – βάρος 30%
(ατομική ή σε ομάδες των 2 ατόμων)

Αντικείμενο της εργασίας

Σκοπός της εργασίας είναι η εμβάθυνση σε ερευνητικά θέματα Διαχείρισης Δεδομένων, μέσα από τη μελέτη άρθρων και τη συγγραφή μιας βιβλιογραφικής εργασίας επισκόπησης (survey paper). Πιο αναλυτικά, θα επιλέξετε μια από τις εργασίες που εμφανίζονται παρακάτω και με αφετηρία αυτή την εργασία θα μελετήσετε το σχετικό χώρο και θα συγγράψετε ένα survey paper. Σκοπός είναι η κατανόηση και συνθετική παρουσίαση μιας ερευνητικής περιοχής (και προφανώς όχι η αποστήθιση-μετάφραση-αντιγραφή υλικού που υπάρχει στο Internet).

- [1] Adadi, A. A survey on data-efficient algorithms in big data era. *Journal of Big Data*, 8, article no. 24 (2021). <https://doi.org/10.1186/s40537-021-00419-9>
- [2] Arundel, S.T., McKeehan, K.G., Campbell, B.B. et al. A guide to creating an effective big data management framework. *J Big Data* 10, 146 (2023). <https://doi.org/10.1186/s40537-023-00801-9>
- [3] Angles, R., Gutiérrez, C. Survey of graph database models. *ACM Computing Surveys*, 40, 1, article no. 1 (2008). <https://doi.org/10.1145/1322432.1322433>
- [4] Dann, J., Ritter, D., Fröning, H., Non-Relational Databases on FPGAs: Survey, Design Decisions, Challenges. *ACM Computing Surveys*, accepted October 2022, (2022). <https://doi.org/10.1145/3568990>
- [5] Dipietro, L., Gonzalez-Mego, P., Ramos-Estebanez, C. et al. The evolution of Big Data in neuroscience and neurology. *J Big Data* 10, 116 (2023). <https://doi.org/10.1186/s40537-023-00751-2>
- [6] Do, T., Graefe, G., Naughton, J., Efficient sorting, duplicate removal, grouping, and aggregation. *ACM Transactions on Database Systems*, accepted September 2022, (2022). <https://doi.org/10.1145/3568027>
- [7] Doukeridis, C., Nørnvåg, K. A survey of large-scale analytical query processing in MapReduce. *The VLDB Journal*, 23, 355–380 (2014). <https://doi.org/10.1007/s00778-013-0319-9>

- [8] Fonseca, J., Bacao, F. Tabular and latent space synthetic data generation: a literature review. *J Big Data* 10, 115 (2023). <https://doi.org/10.1186/s40537-023-00792-7>
- [9] Gul, M., Rehman, M. Big data: an optimized approach for cluster initialization. *J Big Data* 10, 120 (2023). <https://doi.org/10.1186/s40537-023-00798-1>
- [10] Jensen, S. K., Pedersen, T. B., Thomsen, C. Time series management systems: a survey. *IEEE Transactions on Knowledge and Data Engineering*, 29, 2581-2600 (2017). <https://doi.org/10.1109/TKDE.2017.2740932>
- [11] Kong, L., Tan, J., Huang, J., Chen, G., Wang, S., Jin, X., Zeng, P., Khan, M., Edge-Computing-Driven Internet of Things: A Survey. *ACM Computing Surveys*, October 2022, (2022). <https://doi.org/10.1145/3555308>
- [12] Nargesian, F., Pu, K., Ghadhiri-Bashadoost, B., Zhu, E., Miller, R., Data Lake Organization. *IEEE Transactions on Knowledge and Data Engineering*, pp. 237-250, vol. 35, (2023). [10.1109/TKDE.2021.3091101](https://doi.org/10.1109/TKDE.2021.3091101)
- [13] Papadakis, G., Skoutas, D., Thanos, E., Palpanas, T. Blocking and filtering techniques for entity resolution: a survey. *ACM Computing Surveys*, 53, 2, article no. 31 (2021). <https://doi.org/10.1145/3377455>
- [14] Rahul, K., Banyal, R.K. & Arora, N. A systematic review on big data applications and scope for industrial processing and healthcare sectors. *J Big Data* 10, 133 (2023). <https://doi.org/10.1186/s40537-023-00808-2>
- [15] Seifert, M., Kuehnel, S., Sackmann, S., Hybrid Clouds Arising from Software as a Service Adoption: Challenges, Solutions, and Future Research Directions. *ACM Computing Surveys*, accepted October 2022, (2022). <https://doi.org/10.1145/3570156>
- [16] Sim, J., Kim, B., Jeon, K., Joo, M., Lim, J., Lee, J., Choo, K-K. R., Technical Requirements and Approaches in Personal Data Control. *ACM Computing Surveys*, accepted October 2022, (2022). <https://doi.org/10.1145/3558766>
- [17] Vera-Oliveira, H., Guo, R., Huacarpuma, R. C. Data modeling and NoSQL databases – a systematic mapping review. *ACM Computing Surveys*, 54, article no. 116 (2022). <https://doi.org/10.1145/3457608>
- [18] Wylot, M., Hauswirth, M., Cudré-Mauroux, P., Sakr, S. RDF data storage and query processing schemes: a survey. *ACM Computing Surveys*, 51, 4, article no. 84 (2019). <https://doi.org/10.1145/3177850>
- [19] Yuan, G., Lu, J., Yan, Z., Wu, S., A Survey on Mapping Semi-Structured Data and Graph Data to Relational Data. *ACM Computing Surveys*, accepted October 2022, (2022). <https://doi.org/10.1145/3567444>
- [20] Zhang, H., Chen, G., Ooi, B. C., et al. In-memory big data management and processing: a survey. *IEEE Transactions on Knowledge and Data Engineering*, 27, 7, 1920-1948 (2015). <https://doi.org/10.1109/TKDE.2015.2427795>

Όλες οι εργασίες μπορούν να εντοπιστούν μέσα από βιβλιογραφικές πηγές, όπως ACM Digital Library (<https://dl.acm.org>) και DBLP Computer Science Bibliography (<https://dblp.org>), είτε ελεύθερα είτε με πρόσβαση μέσα από το δίκτυο του Παν/μίου. (Σε περίπτωση δυσκολίας να εντοπίσετε το υλικό και για οποιαδήποτε άλλη απορία σχετικά με την εργασία, μπορείτε να απευθυνθείτε στους διδάσκοντες μέσω e-mail.)

Το παραδοτέο της εργασίας θα αποτελείται από το paper (σε source, π.χ. doc ή tex, και pdf) και μία παρουσίαση (σε source, π.χ. ppt, και pdf) που θα συνοψίζει την επισκόπηση που έγινε, σύμφωνα με τα παρακάτω:

- Το paper θα πρέπει να ακολουθεί το πρότυπο μορφοποίησης της ACM για τη συγγραφή ερευνητικών εργασιών, βλ. <https://www.acm.org/publications/taps/word-template-workflow>, και θα έχει έκταση 10-15 σελίδες.
- Η παρουσίαση θα έχει πλήθος διαφανειών ικανό να καλύψει χρόνο 10-12'.

Οδηγίες σχετικά με τη συγγραφή εργασιών τύπου επισκόπησης (survey papers):

You are expected to write a survey paper on a recent research topic related to this class. A survey paper, as described by ACM Computing Surveys, is "a paper that summarizes and organizes recent research results in a novel way that integrates and add understanding to work in the field. A survey article assumes a general knowledge of the area; it emphasizes the classification of the existing literature, developing a perspective on the area, and evaluating trends." For the purpose of this class, "general knowledge of the area" includes the topics we cover in class. You may ignore "evaluating trends".

(πηγή: <https://www.comp.nus.edu.sg/~cs5248/0304S1/survey.html>)

Ενδεικτική δομή ενός survey paper:

- Introduction: Discuss the background; Summarize the surveyed research area and explain why the surveyed area has been studied; Summarize the classification scheme you used to do the survey; Summarize the surveyed techniques with the above classification scheme
- Survey details: Present the surveyed techniques using the classification scheme in detail; Identify the trends in the surveyed area; Give evidence for your decision
- Conclusions and possible future work: Summarize the conclusions of your survey
- References: List all the citations referenced in your paper

Σχετικές πηγές:

- Some thoughts about writing a survey paper (http://www.cs.ucf.edu/~lboloni/Teaching/EEL6788_2008/slides/SurveyTutorial.pdf)
- Writing a survey paper (<http://www.comp.nus.edu.sg/~teoym/cs5239-04/writing-survey-paper.doc>)

Κρίσιμες ημερομηνίες

- Μέχρι τη **Δευτέρα 18/12/23** θα πρέπει να αποστείλετε email προς **npelekis@unipi.gr** με θέμα (subject) του μηνύματος: «CDS110 – Βιβλιογραφική εργασία», όπου θα αναφέρετε τη σύνθεση της ομάδας (ΑΜ, ονοματεπώνυμο, email) και τρία (3) από τα παραπάνω papers που επιθυμείτε να εργαστείτε, κατά σειρά προτίμησης. Η αποστολή αυτού του email είναι υποχρεωτική. Η ανάθεση θα γίνει μέχρι την **Παρασκευή 22/12/22** με απαντητικό email.
- Μέχρι την Κυριακή **14/01/24, 23:55μ.μ.** θα πρέπει να «ανεβάσετε» στο eclass το παραδοτέο της εργασίας σας, όπως έχει περιγραφεί παραπάνω.
- Τη Δευτέρα **15/01/24, 18:00-21:00** θα παρουσιάσετε την εργασία σας στους διδάσκοντες.

Σημείωση: είναι αναπόφευκτο ότι το ίδιο θέμα θα ανατεθεί σε περισσότερες από μια ομάδες. Σε μια τέτοια περίπτωση προφανώς θα εργαστείτε ανεξάρτητα. Σε αντίθετη περίπτωση δυστυχώς θα μηδενιστούν όλες οι εμπλεκόμενες εργασίες.