

**Universidade de São Paulo – USP**  
**Instituto de Ciências Matemáticas e de Computação – ICMC**  
**Pós-Graduação em Ciências de Computação e Matemática Computacional**

**SCC5911 – Procedência de Dados e Data Warehousing**  
**Docente Responsável: Profa. Cristina Dutra de Aguiar Ciferri**

**Objetivos**

Estudo de conceitos sobre procedência dos dados e data warehousing. Investigação de aspectos relacionados à procedência dos dados em processos de integração, à data warehouse espacial, data warehouse espaço-temporal e data warehouse de imagens, e a técnicas para o armazenamento e a indexação de dados nesses ambientes não-convencionais de banco de dados.

**Conteúdo (Ementa)**

1. Procedência dos Dados em Processos de Integração
  - 1.1 Conceitos Básicos
  - 1.2 Quatro Aspectos da Procedência: o que armazenar, como coletar, como armazenar, como consultar
  - 1.3 Integração em Nível de Instância e de Esquema
  - 1.4 Reaplicação de Decisões de Integração
  - 1.5 Modelo de Procedência PrInt
2. Data Warehousing
  - 2.1 Conceitos Básicos
  - 2.2 Data Warehouse Espacial
  - 2.3 Data Warehouse Espaço-Temporal
  - 2.4 Data Warehouse de Imagens
  - 2.5 Indexação em Ambientes de Data Warehousing

**Bibliografia Básica**

- KIMBALL, R., ROSS, M. The Data Warehouse Toolkit: The Complete Guide to Dimensional Modeling, 2nd edition. John Wiley and Sons, Inc., 2002.
- TAN, W.-C. Research problems in data provenance. IEEE Data Engineering Bulletin, v. 27, n. 4, p. 45-52, 2004.
- FREIRE, J.; KOOP, D.; SANTOS, E.; SILVA, C. T. Provenance for computational tasks: A survey. IEEE Computing in Science & Engineering, v. 10, n. 3, p. 11-21, 2008.

## Bibliografia Complementar

AGRAWAL, P., BENJELLOUN, O., SARMA, A. D., HAYWORTH, C., NABAR, S., SUGIHARA, T., WIDOM, J. Trio: a system for data, uncertainty, and lineage. In VLDB '06: Proceedings of the 32nd international conference on Very large data bases, pages 1151–1154. 2006.

ARCHER, D. W., DELCAMBRE, L. M. L., MAIER, D. A framework for fine-grained data integration and curation, with provenance, in a dataspace. In Workshop on the Theory and Practice of Provenance, 2009.

BECKMANN, N.; KRIEGEL, H.-P.; SCHNEIDER, R.; SEEGER, B. The R\*-tree: an efficient and robust access method for points and rectangles. In: Proceedings of the 1990 ACM SIGMOD International Conference on Management of Data, Atlantic City, USA, p. 322–331, 1990.

BECKMANN, N.; SEEGER, B. A revised R\*-tree in comparison with related index structures. In Proceedings of the 2009 ACM SIGMOD International Conference on Management of Data, p. 799-812. 2009.

BENJELLOUN, O., DAS SARMA, A., HALEVY, A., THEOBALD, M., WIDOM, J. Databases with uncertainty and lineage. The VLDB Journal, 17(2):243–264, 2008.

BUNEMAN, P., KHANNA, S., TAN, W. C. Why and where: A characterization of data provenance. In ICDT '01: Proceedings of the 8th International Conference on Database Theory, pages 316–330. 2001.

CARÉLO, C. C. M.; POLA, I. R. V.; CIFERRI, R. R.; TRAINA, A. J. M.; TRAINA-JR, C.; CIFERRI, C. D. A. The Onion-tree: quick indexing of complex data in the main memory. In: Proceedings of the 13th East European Conference on Advances in Databases and Information Systems (ADBIS), Riga, Latvia, p. 1–17, 2009.

CARÉLO, C. C. M.; POLA, I. R. V.; CIFERRI, R. R.; TRAINA, A. J. M.; TRAINA-JR, C.; CIFERRI, C. D. A. Slicing the metric space to provide quick indexing of complex data in the main memory. Information Systems, v. In Press, 2010.

CIFERRI, C. D. A. Distribuição dos Dados em Ambientes de Data Warehousing: O Sistema WebD2W e Algoritmos Voltados à Fragmentação Horizontal dos Dados. Tese de Doutorado, Universidade Federal de Pernambuco (UFPE), 2002.

CIFERRI, R. R. Análise da influência do fator distribuição espacial dos dados no desempenho de métodos de acesso multidimensionais. Tese de Doutorado, Universidade Federal de Pernambuco (UFPE), 2002.

FINKEL, R. A.; BENTLEY, J. L. Quad trees: a data structure for retrieval on composite keys. Acta Informatica, v. 4, p. 1–9, 1974.

FREIRE, J. KOOP, D. SANTOS, E., SILVA, C. T. Provenance for computational tasks: A survey. IEEE Computing Science & Engineering, 10(3):11–21, 2008.

GAEDE, V.; GUNTHER, O. Multidimensional access methods. Journal of Molecular Biology (JMB), v. 30, p. 170–231, 1998.

GUTTMAN, A. R-trees: a dynamic index structure for spatial searching. In: Proceedings of the 1984 ACM SIGMOD International Conference on Management of Data, Boston, USA, p. 47–57, 1984.

IVES Z. G., GREEN T. J., KARVOUNARAKIS G., TAYLOR N. E., TANNEN V., TALUKDAR P. P., JACOB M., PEREIRA F.. The orchestra collaborative data sharing system. SIGMOD Record, 37(3):26–32, 2008.

PAPADOPOULOS, A. N.; CORRAL, A.; NANOPOULOS, A.; THEODORIDIS, Y. R-Tree (and Family). Encyclopedia of Database Systems. 2009.

POLA, I. R. V.; TRAINA-JR, C.; TRAINA, A. J. M. The MM-tree: A memory-based metric tree without overlap between nodes. In: Proceedings of the 12th East European Conference on Advances in Databases and Information Systems (ADBIS), p. 157–171, 2007.

SELLIS, T. K.; ROUSSOPOULOS, N.; FALOUTSOS, C. The R+-tree: a dynamic index for multi-dimensional objects. In: Proceedings of the 13th International Conference on Very Large Data Bases (VLDB), Brighton, England, p. 507–518, 1987.

TOMAZELA, B. MPPI: um modelo de procedência para subsidiar processos de integração. Master's thesis, Universidade de São Paulo, São Carlos, SP, 2010.

TSIROGIANNIS, D.; KOUDAS, N. Suffix tree construction algorithms on modern hardware. In: Proceedings of the EDBT 2010, p. 263-274. 2010.

**+ artigos atuais de periódicos e eventos qualificados**