

## Referências Bibliográficas

- AARTS, E.; LENSTRA, J. **Local search in combinatorial optimization**. 1. ed., New Jersey: Princeton University Press, 2003, 515p.
- ACAMPORA, G.; LOIA, V.; SALERNO, S. ; VITIELLO, A. A hybrid evolutionary approach for solving the ontology alignment problem. **International Journal of Intelligent Systems**, v.27, n.3, p. 189–216, 2012.
- AGRAWAL, R.; AILAMAKI, A.; BERNSTEIN, P.; BREWER, E.; CAREY, M.; CHAUDHURI, S.; DOAN, A.; FLORESCU, D.; FRANKLIN, M.; GARCIA-MOLINA, H. ; OTHERS. The claremont report on database research. **ACM SIGMOD Record**, v.37, n.3, p. 9–19, 2008.
- ALGERGAWY, A.; MASSMANN, S. ; RAHM, E. A clustering-based approach for large-scale ontology matching. **Advances in Databases and Information Systems**, p. 415–428, 2011.
- APACHE. **Jena Documentation Overview**. Apache Software Foundation, 2011. Disponível em <http://incubator.apache.org/jena/documentation/>, último acesso em 21 de janeiro de 2011.
- BAEZA-YATES, R.; RIBEIRO-NETO, B. **Modern Information Retrieve: The Concepts and Technology behind Search**. 2. ed., New York: ACM Press, 2011, 944p.
- BANZHAF, W.; KOZA, J.; RYAN, C.; SPECTOR, L. ; JACOB, C. Genetic programming. **Intelligent Systems and their Applications**, v.15, n.3, p. 74–84, 2000.
- BATITI, R. **Reactive search: Toward self-tuning heuristics**. In: Rayward-Smith, V. J., editor, **MODERN HEURISTIC SEARCH METHODS**, p. 61–83. John Wiley and Sons Ltd, Chichester, UK, 1. ed., 1996.
- BERNSTEIN, P.; RAHM, E. **Data warehouse scenarios for model management**. In: Laender, A. H. F.; Liddle, S. W. ; Storey, V. C., editors, **PROCEEDINGS OF THE 19TH INTERNATIONAL CONFERENCE ON CONCEPTUAL MODELING**, p. 1–15, Heidelberg, 2000. Springer-Verlag.

- BERTINO, E.; CATANIA, B. ; ZARRI, G. **Intelligent database systems**. Addison-Wesley Professional, 2001, 464p.
- BIRATTARI, M.; STÜTZLE, T.; PAQUETE, L. ; VARRENTRAPP, K. **A racing algorithm for configuring metaheuristics**. In: Langdon, W. B., editor, PROCEEDINGS OF THE GENETIC AND EVOLUTIONARY COMPUTATION CONFERENCE (GECCO-2002), volume 2, p. 11–18, San Francisco, CA, USA, 2002. Morgan Kaufmann Publishers.
- BOCK, J.; HETTENHAUSEN, J. Discrete particle swarm optimisation for ontology alignment. **Information Sciences**, v.192, p. 152–173, 2010.
- BOCK, J.; RUDOLPH, S. ; MUTTER, M. More than the sum of its parts – holistic ontology alignment by population-based optimisation. **Foundations of Information and Knowledge Systems**, v.7153, p. 71–90, 2012.
- BONNANS, J. **Numerical optimization: theoretical and practical aspects**. 2. ed., Springer-Verlag New York Inc, 2006, 490p.
- BOUQUET, P.; EUZENAT, J.; FRANCONI, E.; SERAFINI, L.; STAMOU, G. ; TESSARISDE, S. **D2.2.1 specification of a common framework for characterizing alignment**. Technical Report DIT-04-090, Department of Information and Communication Technology, University of Trento, Trento, Italy, 2004.
- BOWER, J.; CHRISTENSEN, C. Disruptive technologies: catching the wave. **Harvard Business Review**, v.73, n.1, p. 43–53, 1995.
- BRANK, J.; GROBELNIK, M. ; MLADENIĆ, D. **A survey of ontology evaluation techniques**. In: PROCEEDINGS OF THE CONFERENCE ON DATA MINING AND DATA WAREHOUSES, número a, p. 166–170, 2005.
- BURANARACH, M. **The foundation for semantic interoperability on the World Wide Web**. 2001. Tese de Doutorado - Department of Information Science and Telecommunications, School of Information Sciences, University of Pittsburgh.
- COSTA, L. A. A. F. **Algoritmos Evolucionários em Optimização Uni e Multi-objetivo**. Braga, Portugal, 2003. 254p. Tese de doutorado - Departamento de Produção e Sistemas, Escola de Engenharia, Universidade do Minho.

- COY, S.; GOLDEN, B.; RUNGER, G. ; WASIL, E. Using experimental design to find effective parameter settings for heuristics. **Journal of Heuristics**, v.7, n.1, p. 77–97, 2001.
- CUEL, R.; DELTEIL, A.; LOUIS, V. ; RIZZI, C. **Knowledge web white paper: The technology roadmap of the semantic web**. Technical report, Knowledge Web Group, 2007. Disponível em <http://knowledgeweb.semanticweb.org/o2i/menu/KWTR-whitepaper-43-final.pdf>, último acesso em 25 de março de 2012.
- DAMERAU, F. A technique for computer detection and correction of spelling errors. **Communications of the ACM**, v.7, n.3, p. 171–176, 1964.
- DO, H.; MELNIK, S. ; RAHM, E. Comparison of schema matching evaluations. **Web, Web-Services, and Database Systems**, v.2593, p. 221–237, 2003.
- DO, H. H. **Schema Matching and Mapping-based Data Integration: Architecture, Approaches and Evaluation**. 1. ed., Saarbrücken, Germany: VDM Verlag, 2006, 228p.
- DUCHATEAU, F.; BELLAHSENE, Z. ; COLETTA, R. **A flexible approach for planning schema matching algorithms**. In: PROCEEDINGS OF THE ON THE MOVE CONFERENCES, volume 1, p. 249–264. Springer-Verlag, 2008.
- DUCHATEAU, F.; COLETTA, R.; BELLAHSENE, Z. ; MILLER, R. **(not) yet another matcher**. In: PROCEEDINGS OF THE ACM CONFERENCE ON INFORMATION AND KNOWLEDGE MANAGEMENT CIKM'09, p. 1537–1540. ACM Press, 2009.
- DUONG, T. H.; NGUYEN, N. T. ; JO, G. **A method for integration of wordnet-based ontologies using distance measures**. In: Lovrek, I.; Howlett, R. J. ; Jain, L. C., editors, PROCEEDINGS OF THE KNOWLEDGE-BASED INTELLIGENT INFORMATION AND ENGINEERING SYSTEMS, volume 5177 de **Lecture Notes in Computer Science**, p. 210–219. Springer, 2008.
- EHRIG, M.; STAAB, S. ; SURE, Y. **Bootstrapping ontology alignment methods with apfel**. In: PROCEEDINGS OF THE INTERNATIONAL SEMANTIC WEB CONFERENCE ISWC'05, p. 186–200. Springer-Verlag, 2005.
- EUZENAT, J. **Towards composing and benchmarking ontology alignments**. In: PROCEEDINGS OF THE ISWC WORKSHOP ON SEMANTIC IN-

FORMATION INTEGRATION, SANIBEL ISLAND (FL US), p. 165–166, Sanibel Island, 2003.

EUZENAT, J. **An api for ontology alignment**. In: PROCEEDINGS OF THE INTERNATIONAL SEMANTIC WEB CONFERENCE, volume 3298 de **Lecture Notes in Computer Science**, p. 698–712, Hiroshima, Japan, 2004. Springer.

EUZENAT, J.; SHVAIKO, P. **Ontology matching**. 1. ed., Springer-Verlag New York Inc, 2007, 333p.

EUZENAT, J.; MEILICKE, C.; STUCKENSCHMIDT, H.; SHVAIKO, P. ; TROJAHN, C. Ontology alignment evaluation initiative: Six years of experience. **Journal on data semantics XV**, v.6720, p. 158–192, 2011.

BREITMAN, K. K.; FELICÍSSIMO, C. H. ; CASANOVA, M. A. **Cato - a lightweight ontology alignment tool**. In: Belo, O.; Eder, J.; e Cunha, J. F. ; Pastor, O., editors, PROCEEDINGS OF THE INTERNATIONAL CONFERENCE ON ADVANCED INFORMATION SYSTEMS ENGINEERING, volume 161 de **CEUR Workshop Proceedings**. CEUR-WS.org, 2005.

FENSEL, D. **Ontologies: a silver bullet for knowledge management and electronic commerce**. 2. ed., Heidelberg: Springer Verlag, 2004.

FENSEL, D.; FACCA, F. M.; SIMPERL, E. ; TOMA, I. **Semantic Web Services**. 1. ed., New York: Springer Verlag, 2011, 255p.

FLETCHER, R. **Practical methods of optimization**. 1. ed., Cambridge University Press, 2004, 730p.

GARTNER. **Gartner Identifies Top Ten Disruptive Technologies for 2008 to 2012**. Gartner, Inc., 2008. Disponível em <http://www.gartner.com/it/page.jsp?id=681107>, último acesso em 19 de março de 2012.

GIUNCHIGLIA, F.; YATSKEVICH, M. ; SHVAIKO, P. Semantic matching: Algorithms and implementation. **Journal on Data Semantics IX**, p. 1–38, 2007.

GOLDBERG, D. **Genetic algorithms in search, optimization, and machine learning**. 1. ed., Addison-Wesley Professional, 1989, 432p.

GOLDBERG, D.; SASTRY, K. **Genetic algorithms: The design of innovation**. 2. ed., Springer-Verlag, 2011, 335p.

- GRACIA, J.; MENA, E. **Web-based measure of semantic relatedness**. In: PROCEEDINGS OF THE WEB INFORMATION SYSTEMS ENGINEERING-WISE 2008, p. 136–150. Springer-Verlag, 2008.
- GUDWIN, R.; VON ZUBEN, F. **Material de aula da disciplina Computação Evolutiva**. Faculdade de Engenharia Elétrica, UNICAMP, 1998. Disponível em <ftp://ftp.dca.fee.unicamp.br/pub/docs/ia707/aula4.pdf>, último acesso em 13 de abril de 2010.
- GÎNSCA, A.-L.; IFTENE, A. **Using a genetic algorithm for optimizing the similarity aggregation step in the process of ontology alignment**. In: PROCEEDINGS OF THE 9TH ROEDUNET INTERNATIONAL CONFERENCE (ROEDUNET), p. 118–122, Sibiu, Romania, 2010. IEEE.
- HAHN, U.; SCHULZ, S.; STAAB, S. ; STUDER, R. **Building a very large ontology from medical thesauri**. p. 133–150. Springer Verlag, 2004.
- HAMMING, R. Error detecting and error correcting codes. **Bell System Technical Journal**, v.29, n.2, p. 147–160, 1950.
- HANSEN, P.; MLADENOVIĆ, N. Variable neighborhood search. **Handbook of metaheuristics**, p. 145–184, 2003.
- HAUPT, R.; HAUPT, S. ; WILEY, J. **Practical genetic algorithms**. 2. ed., Wiley-Interscience, 2004, 272p.
- HEATH, T.; BIZER, C. Linked data: Evolving the web into a global data space. **Synthesis Lectures on the Semantic Web: Theory and Technology**, v.1, n.1, p. 1–136, 2011.
- HOOS, H. H. .; STUTZLE, T. **Empirical analysis of randomized algorithms**. In: Gonzalez, T.F., editor, HANDBOOK OF APPROXIMATION ALGORITHMS AND METAHEURISTICS, chapter 14, p. 14.1–14.17. Chapman and Hall, 1. ed., 2007.
- HUANG, J.; DANG, J.; VIDAL, J. ; HUHNS, M. **Ontology matching using an artificial neural network to learn weights**. In: PROCEEDINGS OF THE IJCAI WORKSHOP ON SEMANTIC WEB FOR COLLABORATIVE KNOWLEDGE ACQUISITION (SWECKA-07), p. 80–85, Hyderabad, India, 2007.
- JARO, M. Advances in record-linkage methodology as applied to matching the 1985 census of tampa, florida. **Journal of the American Statistical Association**, p. 414–420, 1989.

- LANGDON, W.; POLI, R. **Foundations of genetic programming**. 1. ed., Springer-Verlag, 2002, 276p.
- LANZENBERGER, M.; SAMPSON, J. **Human-mediated visual ontology alignment**. In: Smith, M. J.; Salvendy, G., editors, PROCEEDINGS OF THE HUMNA COMPUTER INTERACTION, volume 4558 de **Lecture Notes in Computer Science**, p. 394–403. Springer, 2007.
- LEE, Y.; SAYYADIAN, M.; DOAN, A. ; ROSENTHAL, A. etuner: tuning schema matching software using synthetic scenarios. **The Very Large Database Journal**, v.16, n.1, p. 97–122, 2007.
- LEME, L. A. P. P.; CASANOVA, M. A.; BREITMAN, K. K. ; FURTADO, A. L. **Instance-based owl schema matching**. In: Filipe, J.; Cordeiro, J., editors, PROCEEDINGS OF THE INTERNATIONAL CONFERENCE ON ENTERPRISE INFORMATION SYSTEMS, volume 24 de **Lecture Notes in Business Information Processing**, p. 14–26. Springer, 2009.
- LEME, L. A. P. P.; CASANOVA, M. A.; BREITMAN, K. K. ; FURTADO, A. L. Owl schema matching. **Journal of the Brazilian Computer Society**, v.16, n.1, p. 21–34, 2010.
- LEVENSHTEIN, V. Binary codes capable of correcting deletions, insertions, and reversals. v.10, n.8, p. 707–710, 1966.
- LI, J.; TANG, J.; LI, Y. ; LUO, Q. Rimom: A dynamic multistrategy ontology alignment framework. **IEEE Transactions on Knowledge and Data Engineering**, v.21, n.8, p. 1218–1232, 2009.
- MANNING, C.; RAGHAVAN, P. ; SCHUTZE, H. **Introduction to information retrieval**. 1. ed., Cambridge University Press, 2008, 496p.
- MAO, M.; PENG, Y. ; SPRING, M. **Neural network based constraint satisfaction in ontology mapping**. In: PROCEEDINGS OF THE CONFERENCE ON ARTIFICIAL INTELLIGENCE AAAI '08, p. 1207–1212. AAAI Press, 2008.
- MARTINEZ-GIL, J.; NAVAS-DELGADO, I. ; ALDANA-MONTES, J. **Sifo. an efficient taxonomical matcher for ontology alignment**. Technical Report ITI-08-3, Department of Languages and Computing Sciences, University of Malaga, 2008.

- MARTINEZ-GIL, J.; ALDANA-MONTES, J. Evaluation of two heuristic approaches to solve the ontology meta-matching problem. **Knowledge and Information Systems**, v.26, n.2, p. 225–247, 2009.
- MARTINEZ-GIL, J.; ALDANA-MONTES, J. F. An overview of current ontology meta-matching solutions. **Knowledge Engineering Review**, 2012, a ser publicado.
- MATUSZEK, C.; CABRAL, J.; WITBROCK, M. ; DE OLIVEIRA, J. **An introduction to the syntax and content of cyc**. In: PROCEEDINGS OF THE 2006 AAAI SPRING SYMPOSIUM ON FORMALIZING AND COMPILING BACKGROUND KNOWLEDGE AND ITS APPLICATIONS TO KNOWLEDGE REPRESENTATION AND QUESTION ANSWERING, p. 44–49, 2006.
- MITCHELL, M. **An introduction to genetic algorithms**. 1. ed., Cambridge, Massachusetts: The MIT Press, 1998, 211p.
- MOSCATO, P.; COTTA, C. Una introducción a los algoritmos meméticos. **Revista Iberoamericana de Inteligencia Artificial**, v.7, n.19, p. 131–148, 2003.
- NEBRO, A.; LUNA, F.; ALBA, E.; DORRONSORO, B.; DURILLO, J. ; BEHAM, A. Abyss: Adapting scatter search to multiobjective optimization. **IEEE Transactions on Evolutionary Computation**, v.12, n.4, p. 439–457, 2008.
- OAEI. **Benchmark test library: test at a glance**. Ontology Alignment Evaluation Initiative, 2011. Disponível em <http://oaei.ontologymatching.org/2011/benchmarks/>, último acesso em 19 de fevereiro de 2012.
- PACHECO, M. Algoritmos genéticos: princípios e aplicações. **ICA: Laboratório de Inteligência Computacional Aplicada, Departamento de Engenharia Elétrica, Pontifícia Universidade Católica do Rio de Janeiro**, 1999. Disponível em <http://www.mendeley.com/research/algoritmos-geneticos-principios-e-aplicaes-1/>, último acesso em 25 de novembro de 2010.
- POLI, R.; LANGDON, W. ; MCPHEE, N. **A field guide to genetic programming**. 1. ed., San Francisco, California: Lulu Enterprises Uk Ltd, 2008, 235p.

- RAHM, E.; BERNSTEIN, P. A survey of approaches to automatic schema matching. **The Very Large Database Journal**, v.10, n.4, p. 334–350, 2001.
- RAHM, E. Towards large-scale schema and ontology matching. **Schema Matching and Mapping**, p. 3–27, 2011.
- ROBERTSON, S.; JONES, K. Relevance weighting of search terms. **Journal of the American Society for Information Science**, v.27, n.3, p. 129–146, 1976.
- SANTOS, C. T.; EUZENAT, J.; TAMMA, V. ; PAYNE, T. **Argumentation for reconciling agent ontologies**. In: Eli, A.; Kon, M. ; Orgun, M., editors, SEMANTIC AGENT SYSTEMS, volume 1, chapter 5, p. 89—111. Springer, 2011.
- SHI, Y.; EBERHART, R. **A modified particle swarm optimizer**. In: PROCEEDINGS OF THE IEEE INTERNATIONAL CONFERENCE ON EVOLUTIONARY COMPUTATION, p. 69–73. IEEE Press, 1998.
- SHVAIKO, P.; EUZENAT, J. A survey of schema-based matching approaches. **Journal on Data Semantics IV**, p. 146–171, 2005.
- SHVAIKO, P.; EUZENAT, J. **Ten challenges for ontology matching**. In: PROCEEDINGS OF THE 7TH INTERNATIONAL CONFERENCE ON ONTOLOGIES, DATABASES, AND APPLICATIONS OF SEMANTICS (ODBASE), p. 1164–1182, Monterrey (MX), 2008. Springer.
- SHVAIKO, P.; EUZENAT, J. Ontology matching: state of the art and future challenges. **IEEE Transactions on Knowledge and Data Engineering**, v.1, n.99, p. 1–20, 2012.
- SOUZA, J. M. **Software Tools for Conceptual Schema Integration**. 1986. Tese de doutorado - University of East Anglia.
- SOUZA, J. F.; SIQUEIRA, S. W. M.; AZEVEDO, L.; BAIÃO, F. ; LOPES, M. **Gestão de ontologias**. Technical Report 0002, Departamento de Informática Aplicada, UNIRIO, 2008.
- SOUZA, J. F.; SIQUEIRA, S. W. M. ; MELO, R. N. **Framework para negociação de significados em sistemas multiagente**. In: PROCEEDINGS OF THE IADIS CONFERÊNCIA IBERO-AMERICANA WWW/INTERNET (CIAWI'09), p. 3–10, Alcalá, Espanha, 2009. IADIS - International Association for Development of the Information Society.



- SOUZA, J. F.; MELO, R. N.; OLIVEIRA, J. ; SOUZA, J. M. **Combining resemblance functions for ontology alignment**. In: Kotsis, G.; Taniar, D.; Pardede, E. ; Ibrahim, I. K., editors, PROCEEDINGS OF THE INTERNATIONAL CONFERENCE ON INFORMATION INTEGRATION AND WEB-BASED APPLICATIONS & SERVICES (IIWAS2009), p. 64–71, Kuala Lumpur, Malaysia, 2009. ACM Press.
- SOUZA, J. F.; SIQUEIRA, S. W. M. ; MELO, R. N. **Adding meaning negotiation skills in multiagent systems**. In: PROCEEDINGS OF THE IEEE INTERNATIONAL CONFERENCE ON INTELLIGENT COMPUTING AND INTELLIGENT SYSTEMS (ICIS 2009), p. 663–667, Shanghai, 2009. IEEE Press.
- SOUZA, J. F.; SIQUEIRA, S. W. M. ; MELO, R. N. **Evolution in ontology-based user modeling**. In: Lytras, M. D.; de Pablos, P. O. ; Damiani, E., editors, SEMANTIC WEB PERSONALIZATION AND CONTEXT AWARENESS: MANAGEMENT OF PERSONAL IDENTITIES AND SOCIAL NETWORKING. IGI Global, Hershey, Pennsylvania, 2010.
- SOUZA, J. F.; MELO, R. N.; OLIVEIRA, J.; SOUZA, J. M. ; SIQUEIRA, S. W. M. Improving software agent communication with structural ontology alignment methods. **International Journal of Information Technology and Web Engineering**, v.5, p. 49–64, 2010.
- SOUZA, J. F.; MELO, R. N.; OLIVEIRA, J. ; SOUZA, J. M. Uma abordagem estrutural para calcular similaridade entre conceitos de ontologias. **Revista de Informática Teória e Aplicada**, v.17, n.2, p. 249–269, 2011.
- SOUZA, J. F.; SIQUEIRA, S. W. M. ; MELO, R. N. **Providing information from brazilian politicians using linked data**. In: de Pablos, P. O.; Lytras, M. D.; Tennyson, R. ; Gayo, J. E. L., editors, CASES ON OPEN-LINKED DATA AND SEMANTIC WEB APPLICATIONS. IGI Global, Hershey, Pennsylvania, 2012. Em fase de editoração.
- SOUZA, J. F.; SIQUEIRA, S. W. M. ; MELO, R. N. **Applying ontology similarity functions to improve software agent communication**. In: MODELS FOR CAPITALIZING ON WEB ENGINEERING ADVANCEMENTS: TRENDS AND DISCOVERIES, p. 43–57. IGI Global, 2012.
- SPOHR, D.; HOLLINK, L. ; CIMIANO, P. **A machine learning approach to multilingual and cross-lingual ontology matching**. In: PROCEEDINGS OF THE 10TH INTERNATIONAL SEMANTIC WEB CONFERENCE (ISWC), p. 665–680, 2011.

- STAAB, S.; STUCKENSCHMIDT, H. **Semantic Web and Peer-to-peer: Decentralized Management and Exchange of Knowledge and Information**. 1. ed., New York: Springer Verlag, 2006, 366p.
- STOFFEL, K.; TAYLOR, M. ; HENDLER, J. **Efficient management of very large ontologies**. In: PROCEEDINGS OF THE NATIONAL CONFERENCE ON ARTIFICIAL INTELLIGENCE, p. 442–447. John Wiley & Sons LTD, 1997.
- STOJANOVIC, L.; MAEDCHE, A.; MOTIK, B. ; STOJANOVIC, N. **User-driven ontology evolution management**. In: PROCEEDINGS OF THE 13TH INTERNATIONAL CONFERENCE ON KNOWLEDGE ENGINEERING AND KNOWLEDGE MANAGEMENT (EKAW-02), volume 2473 de **Lecture Notes in Computer Science**, p. 285–300. Springer, 2002.
- STOILLOS, G.; STAMOU, G. ; KOLLIAS, S. **A string metric for ontology alignment**. In: PROCEEDINGS OF THE INTERNATIONAL SEMANTIC WEB CONFERENCE, p. 624–637. Springer, 2005.
- SUMIDA, B.; HOUSTON, A.; MCNAMARA, J. ; HAMILTON, W. Genetic algorithms and evolution. **Journal of Theoretical Biology**, v.147, n.1, p. 59–84, 1990.
- SYCARA, K.; PAOLUCCI, M.; ANKOLEKAR, A. ; SRINIVASAN, N. Automated discovery, interaction and composition of semantic web services. **Web Semantics: Science, Services and Agents on the World Wide Web**, v.1, n.1, p. 27–46, 2003.
- TALBI, E.-G. **Metaheuristics: from design to implementation**. 1. ed., Hoboken, New Jersey: John Wiley & Sons Inc, 2009, 594p.
- TANOMARU, J. **Motivação, fundamentos e aplicações de algoritmos genéticos**. In: II CONGRESSO BRASILEIRO DE REDES NEURAI, p. 373–403, 1995.
- UKKONEN, E. Approximate string-matching with q-grams and maximal matches. **Theoretical Computer Science**, v.92, n.1, p. 191–211, 1992.
- VIEIRA, C. E. C. **Heurísticas para o problema das p-Mediana Conectadas**. Rio de Janeiro, 2006. 191p. Tese de Doutorado - Pós-graduação em Informática, Pontifícia Universidade Católica do Rio de Janeiro.

- VOLZ, J.; BIZER, C.; GAEDKE, M. ; KOBILAROV, G. **Silk—a link discovery framework for the web of data**. In: PROCEEDINGS OF THE 2ND LINKED DATA ON THE WEB WORKSHOP, p. 1–6, 2009.
- WANG, J.; DING, Z. ; JIANG, C. **Gaom: Genetic algorithm based ontology matching**. In: PROCEEDINGS OF THE IEEE ASIA-PACIFIC CONFERENCE ON SERVICES COMPUTING, p. 617–620, Guangzhou, Guangdong, 2006. IEEE Press.
- WHITLEY, D. An overview of evolutionary algorithms: practical issues and common pitfalls. **Information and Software Technology**, v.43, n.14, p. 817–831, 2001.
- WIESMAN, F.; ROOS, N. ; VOGT, P. **Automatic ontology mapping for agent communication**. In: PROCEEDINGS OF THE FIRST INTERNATIONAL JOINT CONFERENCE ON AUTONOMOUS AGENTS AND MULTI-AGENT SYSTEMS: PART 2, p. 563–564. ACM Press, 2002.
- WINKLER, W. E. **The state of record linkage and current research problems**. Technical Report 99/04, Statistical Research Division, US Census Bureau, 1999.
- ZAIHRAYEU, I. **Towards Peer-to-Peer Information Management Systems**. Trento, Italy, 2006. 164p. Tese de Doutorado - International Doctorate School in Information and Communication Technologies, University of Trento. Disponível em <http://disi.unitn.it/ilya/Download/Publications/PhD-Thesis.pdf>, último acesso em 25 de fevereiro de 2011.

## A

### Algoritmos para medidas de similaridade de texto

Este apêndice apresenta os algoritmos disponibilizados no protótipo do GNoSIS+ para calcular similaridade entre duas cadeias de caracteres. Todos são algoritmos clássicos da literatura de teoria da informação.

Algoritmos para medidas de similaridade de texto são funções que recebem duas cadeias de caracteres e retornam um valor real, geralmente entre 0 e 1, que representa o quão próximo um conjunto de caracteres está do outro. Uma heurística muito adotada na literatura de teoria da informação é que, quanto mais próximos são dois conjuntos de caracteres, maior a similaridade entre eles (Brank et al, 2005).

**Definição A.1 (Similaridade de texto)** *Uma similaridade de texto é uma similaridade  $\delta : \mathbb{S} \times \mathbb{S} \rightarrow [0 \ 1] \cap \mathfrak{R}$ , tal que  $\delta = 1$  representa textos muito similares,  $\delta = 0$  representa textos nada similares e valores entre 0 e 1 representam o grau de certeza na similaridade entre os dois textos.*

As seções seguintes definem as similaridade de texto implementadas no GNoSIS+.

#### A.1

##### Igualdade

Uma implementação simples de igualdade de textos pode ser encontrada classe *EqualsEditDistance*.

**Definição A.2 (Igualdade de textos)** *Uma igualdade de textos é uma similaridade  $\delta : \mathbb{S} \times \mathbb{S} \rightarrow [0 \ 1] \cap \mathfrak{R}$ , tal que  $\forall x, y \in \mathbb{S}, \delta(x, x) = \delta(\bar{x}, \bar{x}) = 1$  e  $\delta(x, y) = \delta(\bar{x}, \bar{y}) = 0$  se  $\bar{x} \neq \bar{y}$ , onde  $\bar{s}$  denota a cadeia de caracteres resultante da normalização de  $s$ .*

Neste trabalho, é considerado como normalização de um texto o texto resultado da substituição dos caracteres maiúsculos por minúsculos e a eliminação de sinais de pontuação e de espaços em branco no início e fim do texto.

## A.2

### Hamming

A distância de Hamming (Hamming, 1950) é dada pelo número substituições de caracteres necessários para transformar uma cadeia de caracteres em outra. Geralmente o algoritmo de Hamming é utilizado para textos de mesmo tamanho. Contudo, foi implementado a versão normalizada pelo comprimento do maior texto, modificada de (Euzenat & Shvaiko, 2007).

**Definição A.3 (Distância de Hamming)** *Sejam dois textos  $s$  e  $t$ , a distância de Hamming é uma similaridade  $\delta : \mathbb{S} \times \mathbb{S} \rightarrow [0, 1]$ , tal que:*

$$\delta(s, t) = 1 - \frac{(\sum_{i=1}^{\min(|s|, |t|)} s[i] \neq t[i]) + ||s| - |t||}{\max(|s|, |t|)} \quad (\text{A-1})$$

Na versão implementada, é realizado um processo de normalização antes da execução do algoritmo. A implementação dessa distância está disponível na classe *HammingEditDistance*.

## A.3

### Levenshtein

A distância de Levenshtein (Levenshtein, 1966) é dada pelo número mínimo de operações necessárias para transformar uma cadeia de caracteres em outra através das operações de inserção, eliminação ou substituição de um caractere.

**Definição A.4 (Distância de Levenshtein)** *Sejam  $s$  e  $t$  duas uma sequências de caracteres  $s = s_1 \dots s_k$  e  $t = t_1 \dots t_l$  são definidas as seguintes operações:*

- uma **deleção**: a transformação  $s_1 \dots s_i \dots s_k \rightarrow s_1 \dots \check{s}_i \dots s_k$  para algum  $1 \leq i \leq k$ , onde  $\check{s}_i$  denota a omissão o  $i$ -ésimo caractere.
- uma **inserção**: a transformação  $s_1 \dots s_i \dots s_k \rightarrow s_1 \dots s_i x \dots s_k$  para algum  $1 \leq i \leq k$  e algum caractere  $x \in \Sigma$ .
- uma **substituição**: a transformação  $s_1 \dots s_i \dots s_k \rightarrow s_1 \dots s_{i-1} x s_{i+1} \dots s_k$  para algum  $1 \leq i \leq k$  e algum caractere  $x \in \Sigma$ .

A distância de Levenshtein é dada por:

$$\delta(s, t) = 1 - \frac{\min(\Gamma^{s,t})}{\max(|s|, |t|)}, \quad (\text{A-2})$$

onde  $\Gamma^{s,t}$  representa a sequência de operações de deleção, inserção e substituição de tal forma que  $s = s_1 \dots s_k \rightarrow t_1 \dots t_l$ .

A implementação dessa distância está disponível na classe *LevenshteinEditDistance*.

#### A.4 Jaro-Winkler

O cálculo de Jaro foi definido para correspondências de nomes próprios que pode conter erros de ortografia similares. A medida Jaro é dada pela soma ponderada da porcentagem de caracteres correspondentes de cada texto e de caracteres transpostos (Jaro, 1989). Winkler (1999) alterou esta medida para combinar textos com longos prefixos comuns e depois redimensionado-la utilizando uma função definida em trechos, cuja periodicidade e pesos dependem do tipo de cadeia de caractere (nome, sobrenome, rua, etc.)

**Definição A.5 (Distância de Jaro-Winkler)** *A distância de Jaro-Winkler é uma medida não simétrica  $\delta : \mathbb{S} \times \mathbb{S} \rightarrow [0, 1]$ , definida como:*

$$\delta(s, t) = 1 - (\delta_{Jaro}(s, t) + P \times Q \times \frac{(1 - \delta_{Jaro}(s, t))}{10}), \quad (\text{A-3})$$

de tal forma que  $P$  é o tamanho do prefixo comum,  $Q$  é uma constante e  $\delta_{Jaro}$  é dado por:

$$\delta_{Jaro}(s, t) = \frac{1}{3} \left( \frac{com(s, t)}{|s|} + \frac{com(t, s)}{|t|} + \frac{com(s, t) - |transp(s, t)|}{|com(s, t)|} \right), \quad (\text{A-4})$$

sendo  $s[i] \in com(s, t)$  se e somente se  $\exists j \in [i - (\min(|s|, |t|)/2, i) + \min(|s|, |t|)/2]$  e  $transp(s, t)$  são os elementos de  $com(s, t)$  que ocorrem em uma ordem diferente em  $s$  e  $t$ .

A implementação dessa medida está disponível na classe *JaroWinklerEditDistance*.

#### A.5 Damerau-Levenshtein

A distância de Damerau-Levenshtein (Damerau, 1964) é dada pelo número mínimo de operações necessárias para transformar um texto em outro através de operações de inserção, eliminação, substituição de um caractere e a transposição de dois caracteres adjacentes, o que diferencia da distância de Levenshtein.

**Definição A.6 (Distância de Damerau-Levenshtein)** *Sejam  $s$  e  $t$  duas uma sequências de caracteres  $s = s_1 \dots s_k$  e  $t = t_1 \dots t_l$  são definidas as seguintes operações:*

- uma **deleção**: a transformação  $s_1 \dots s_i \dots s_k \rightarrow s_1 \dots \check{s}_i \dots s_k$  para algum  $1 \leq i \leq k$ , onde  $\check{s}_i$  denota a omissão o  $i$ -ésimo caractere.
- uma **inserção**: a transformação  $s_1 \dots s_i \dots s_k \rightarrow s_1 \dots s_i x \dots s_k$  para algum  $1 \leq i \leq k$  e algum caractere  $x \in \Sigma$ .
- uma **substituição**: a transformação  $s_1 \dots s_i \dots s_k \rightarrow s_1 \dots s_{i-1} x s_{i+1} \dots s_k$  para algum  $1 \leq i \leq k$  e algum caractere  $x \in \Sigma$ .
- uma **transposição**: a transformação  $s_1 \dots s_i s_{i+1} \dots s_k \rightarrow s_1 \dots s_{i+1} s_i \dots s_k$  para algum  $1 \leq i \leq k$ .

A distância de Damerau-Levenshtein é dada por:

$$\delta(s, t) = 1 - \frac{\min(\Gamma^{s,t})}{\max(|s|, |t|)}, \quad (\text{A-5})$$

onde  $\Gamma^{s,t}$  representa a sequência de operações de deleção, inserção, substituição e transposição de tal forma que  $s = s_1 \dots s_k \rightarrow t_1 \dots t_l$ .

A implementação dessa medida está disponível na classe *DamerauLevenshteinEditDistance*.

## A.6 TFIDF

O TFIDF (*Term Frequency–Inverse Document Frequency*) é um número estatístico que reflete o quão importante uma palavra é para um documento numa coleção (Robertson & Jones, 1976).

Esta é uma medida utilizada na área de Recuperação da Informação (Manning et al, 2008) e implementada neste trabalho como uma medida de similaridade aplicada para avaliar a relevância de uma *substring* em uma cadeia de caracteres ao comparar a frequência de ocorrências da cadeia de caracteres numa entidade da ontologia em relação a sua frequência em todas as entidades.

A medida pode ser definida, como abaixo, modificado de (Euzenat & Shvaiko, 2007):

**Definição A.7 (TFIDF)** Dado um corpus  $C$  de multiconjuntos, são definidas as seguintes medidas:

$$\begin{aligned} \forall t \in \mathbb{S}, \forall s \in C, tf(t, s) &= t\#s && (\text{frequência do termo}) \\ \forall t \in \mathbb{S}, idf(t) &= \log\left(\frac{|C|}{|\{s \in C; t \in s\}|}\right) && (\text{frequência inversa do documento}) \\ TFIDF(s, t) &= tf(t, s) \times idf(t) && (TFIDF) \end{aligned}$$

Sendo que  $t\#s$  é o número de ocorrências de  $t$  em  $s$ .

A implementação dessa medida está disponível na classe *TokenizerEditDistance*.

## B

### Arquivos XML utilizados para configuração do sistema

A listagem abaixo apresenta o arquivo XML utilizado para a avaliação do sistema.

```

1 <?xml version="1.0" encoding="UTF-8"?>
2 <similarityCalc>
3   <container name="FuncoesPrincipais">
4     <function weight="0.3">
5       <class>br.ufjf.ontology.gnosis.similarity.structure.ConceptNameSimilarity</class>
6       <strategy>br.ufjf.ontology.gnosis.similarity.editdistance.LevenshteinEditDistance</
       strategy>
7     </function>
8     <function weight="0.3">
9       <class>br.ufjf.ontology.gnosis.similarity.structure.DirectIndividualSimilarity</
       class>
10      <combination>br.ufjf.ontology.gnosis.similarity.combination.FirstMatchCombination</
       combination>
11      <strategy>br.ufjf.ontology.gnosis.similarity.editdistance.LevenshteinEditDistance</
       strategy>
12      <penalty>0.1</penalty>
13    </function>
14    <function weight="0.0">
15      <class>br.ufjf.ontology.gnosis.similarity.structure.CommentTokenSimilarity</class>
16      <strategy>br.ufjf.ontology.gnosis.similarity.editdistance.TokenizerEditDistance</
       strategy>
17    </function>
18    <container name="FuncoesHierarquia" weight="0.2">
19      <function weight="0.5">
20        <class>br.ufjf.ontology.gnosis.similarity.structure.DirectSuperClassSimilarity</
       class>
21        <combination>br.ufjf.ontology.gnosis.similarity.combination.
       FirstMatchCombination</combination>
22        <strategy>br.ufjf.ontology.gnosis.similarity.editdistance.
       LevenshteinEditDistance</strategy>
23        <penalty>0.1</penalty>
24      </function>
25      <function weight="0.5">
26        <class>br.ufjf.ontology.gnosis.similarity.structure.DirectSubClassSimilarity</
       class>
27        <combination>br.ufjf.ontology.gnosis.similarity.combination.
       FirstMatchCombination</combination>
28        <strategy>br.ufjf.ontology.gnosis.similarity.editdistance.
       LevenshteinEditDistance</strategy>
29        <penalty>0.0</penalty>
30      </function>
31    </container>
32    <container name="FuncoesPropriedades" weight="0.2">
33      <function weight="0.6">
34        <class>br.ufjf.ontology.gnosis.similarity.structure.
       DirectPropertybyRangeDomainSimilarity</class>
35        <combination>br.ufjf.ontology.gnosis.similarity.combination.
       FirstMatchCombination</combination>
36        <strategy>br.ufjf.ontology.gnosis.similarity.editdistance.
       LevenshteinEditDistance</strategy>
37        <penalty>0.1</penalty>
38      </function>
39      <function weight="0.4">
40        <class>br.ufjf.ontology.gnosis.similarity.structure.
       DirectPropertybyNameSimilarity</class>

```



```

41     <combination>br.ufjf.ontology.gnosis.similarity.combination.
        FirstMatchCombination</combination>
42     <strategy>br.ufjf.ontology.gnosis.similarity.editdistance.
        LevenshteinEditDistance</strategy>
43     <penalty>0.1</penalty>
44 </function>
45 </container>
46 </container>
47 <ontologies>
48 <ontology id="/home/benchmarks/101/onto.rdf"></ontology>
49 <ontology id="/home/benchmarks/238/onto.rdf"></ontology>
50 <!-- The weights above will not be used by application if pre-alignment element is
        filled-->
51 <pre-alignment>home/benchmark/settings/prealign238.rdf</pre-alignment>
52 </ontologies>
53 </similarityCalc>

```

Cada teste realizado continha como entrada um arquivo de pré-alinhamento de nome prealignXXX.rdf contendo os alinhamentos conhecidos. Esses alinhamentos são utilizados para alimentar o algoritmo genético para calibragem das funções de similaridade. A lista abaixo exige um arquivo de pré-alinhamento utilizado.

```

1 <?xml version="1.0" encoding="iso-8859-1"?>
2 <rdf:RDF xmlns="http://knowledgeweb.semanticweb.org/heterogeneity/alignment" xmlns:rdf="
    http://www.w3.org/1999/02/22-rdf-syntax-ns#" xmlns:xsd="http://www.w3.org/2001/
    XMLElementSchema#">
3 <Alignment>
4 <xml:yes</xml>
5 <level>0</level>
6 <type>11</type>
7 <onto1>http://oaei.ontologymatching.org/2007/benchmarks/101/onto.rdf</onto1>
8 <onto2>http://oaei.ontologymatching.org/2007/benchmarks/248/onto.rdf</onto2>
9 <uri1>http://oaei.ontologymatching.org/2007/benchmarks/101/onto.rdf</uri1>
10 <uri2>http://oaei.ontologymatching.org/2007/benchmarks/248/onto.rdf</uri2>
11 <map>
12 <Cell>
13 <entity1 rdf:resource="http://oaei.ontologymatching.org/2007/benchmarks/101/onto.rdf
    #PersonList"/>
14 <entity2 rdf:resource="http://oaei.ontologymatching.org/2007/benchmarks/248/onto.rdf
    #dsqdbz"/>
15 <measure rdf:datatype="http://www.w3.org/2001/XMLSchema#float">1.0</measure>
16 <relation>=</relation>
17 </Cell>
18 </map>
19 <map>
20 <Cell>
21 <entity1 rdf:resource="http://oaei.ontologymatching.org/2007/benchmarks/101/onto.rdf
    #Unpublished"/>
22 <entity2 rdf:resource="http://oaei.ontologymatching.org/2007/benchmarks/248/onto.rdf
    #zeadza"/>
23 <measure rdf:datatype="http://www.w3.org/2001/XMLSchema#float">1.0</measure>
24 <relation>=</relation>
25 </Cell>
26 </map>
27 <map>
28 <Cell>
29 <entity1 rdf:resource="http://oaei.ontologymatching.org/2007/benchmarks/101/onto.rdf
    #key"/>
30 <entity2 rdf:resource="http://oaei.ontologymatching.org/2007/benchmarks/248/onto.rdf
    #zdzbh"/>
31 <measure rdf:datatype="http://www.w3.org/2001/XMLSchema#float">1.0</measure>
32 <relation>=</relation>
33 </Cell>
34 </map>
35 </Alignment>
36 </rdf:RDF>

```

O módulo de calibragem dos pesos para as funções de similaridade é configurado através de um arquivo de configuração. A listagem abaixo

apresenta o estado do arquivo de configuração utilizado na avaliação do sistema.

```
#####  
# VALORES INICIAIS  
#####  
ag.geracoes = 5  
ag.populacao.inicial = 300  
# A granularidade dita qual sera a granularidade de discretizacao do problema  
ag.granularidade = 0.025  
ag.buscalocal.periodicidade = 5  
# 1 para 100%, 2 para 50%, 3 para 1/3, 4 para 25% ,etc. 0 para nao inserir nenhuma solucao  
alternativa  
ag.buscalocal.insercao = 0  
# 0 para comparacao difTotal e 1 para comparacao com cada equacao  
ag.tipo.comparacao = 1  
  
#####  
# RENOVACAO DA POPULACAO  
#####  
ag.porcentagem.selecao = 0.5  
ag.porcentagem.individuos.otimos = 0.3  
ag.porcentagem.individuos.ruins = 0.2  
# geracao em que o individuo morre  
ag.mortalidade = 5  
  
#####  
# TAXAS  
#####  
ag.taxa.cruzamento = 0.8  
ag.taxa.mutacao = 0.1  
  
#####  
# DEBUG  
#####  
# 0 para INFO basica, 1 para INFO completa, 2 para DEBUG  
ag.debug.level = 1  
# 1 para criar arquivo de log, 0 para nao criar o arquivo  
ag.debug.log = 0
```