Domain labeling in the Morais dictionary:
bringing structure to unstructured lexicographic data

Ana Salgado, Rute Costa, Toma Tasovac, Fahad Khan,
Bruno Almeida, Margarida Ramos, Sara Carvalho,
Raquel Silva, Mohamed Khemakhem, Laurent Romary

DSNA-24: Dictionary Society of North America 24th Biennial Conference

1 June 2023

CLUNL
CENTRO DE LINGUÍSTICA DA UNIVERSIDADE NOVA DE LISBOA

NOVAFCSH
FACULDADE DE CIÊNCIAS SOCIAIS E HUMANAS
UNIVERSIDADE NOVA DE LISBOA

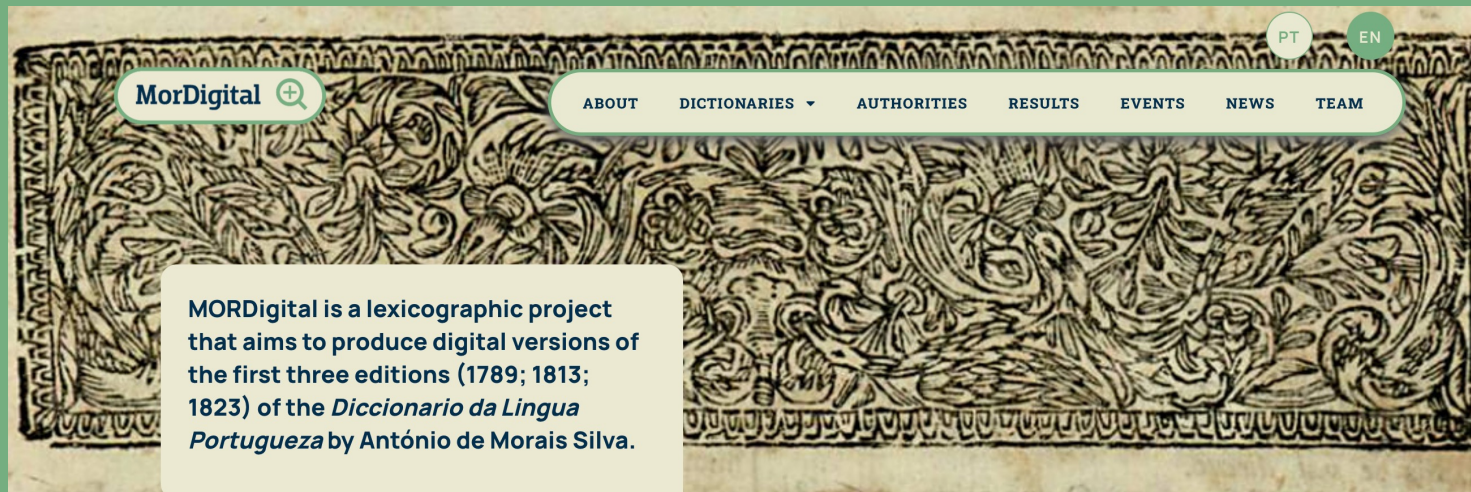fct Fundação
para a Ciência
e a Tecnologia

Outline

- Theoretical background
- MorDigital project
  - The lexicographer
  - The dictionary
  - The editions
- Domain labeling
- Knowledge organization
- Concluding remarks

Theoretical background

- Lexicography and terminology as autonomous disciplines can be found in the broad field of digital humanities.

- The digital revolution has imposed on lexicography the challenge of adopting new methods concerning the traditional research methodology.

- Our approach assumes that terminology and lexicography can be complementary – perceiving lexicography, terminology, ontologies and computational linguistics as an integral part of the digital humanities necessarily implies a paradigm shift in constructing dictionary resources.

- The project draws contributions from multiple domains, from lexicography to terminology, ontologies, linked data and digital humanities, embodying a new paradigm for lexicography.

- The diversity of incompatible formats in the European lexicographic scenario creates severe problems in the digital landscape, making it impossible to interconnect resources and their respective metadata.
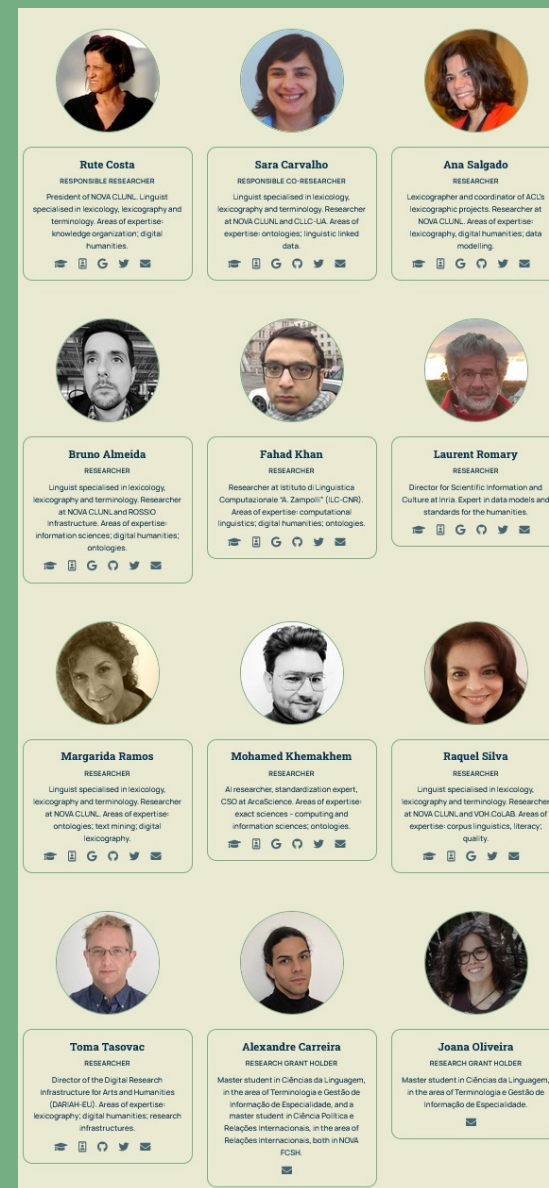
MorDigital project

- This work is part of an ongoing Portuguese national linguistic project (MORDigital – *Digitization of Diccionario da Lingua Portugueza* by António de Morais Silva [PTDC/LLT-LIN/6841/2020], funded by the Portuguese National Funding agency through the FCT (Fundação para a Ciência e Tecnologia).

- This project is not merely of a lexicographic nature, but it also explores the convergence between lexicography and other research domains, such as terminology, ontologies, linked data, and digital humanities.

MORDigital is a lexicographic project that aims to produce digital versions of the first three editions (1789; 1813; 1823) of the *Diccionario da Lingua Portugueza* by António de Morais Silva.

https://mordigital.fcsh.unl.pt/

MorDigital project

- Portuguese national project
  - It includes collaborators from all over Europe:

    - NOVA CLUNL – Centro de Linguística da Universidade NOVA de Lisboa, Portugal
    - Academia das Ciências de Lisboa, Portugal
    - Istituto Di Linguistica Computazionale 'A. Zampolli', Italy
    - CLLC – Centro de Línguas, Literaturas e Culturas da Universidade de Aveiro, Portugal
    - Inria, ALMAnaCH team, France
    - ArcaScience, France
    - BCDH – Belgrade Center for Digital Humanities, Serbia

Salgado et al. (2023). Domain labelling in the Morais dictionary

MorDigital project

- Has two main goals:

  - to convert the first three editions (1789; 1813; 1823) of Morais into a structured lexical resource in both TEI-XML (based on the ISO LMF standard) and in RDF (based on the OntoLex-Lemon model and its recent extensions) to make them available online via TEI Publisher on the website (at the moment only PDFs are available);

  - to describe the lexicographic components of these editions following a rigorous linguistic treatment.

MorDigital project

MorDigital ⊕

Digitisation of dictionary sources

Re-OCRisation of public domain versions of the Morais dictionary. We are using ABBYY FineReader to correct OCR problems.

Structuring of digitized sources

Using GROBID-Dictionaries, a machine learning library for structuring digitized lexical resources and entry-based documents with encyclopedic or bibliographic content. It allows the parsing, extraction and structuring of lexical information.

Conversion and mapping of the TEI content to the LMF standard and their respective serializations, as well as to OntoLex-Lemon.

Labels ontologies and KOS

Knowledge organization and the possibility of building ontologies.

Linked data conversion, annotation and alignment

We aim to link the data and align the entries, senses and other lexicographic content between the three editions.

Platform development

Conception of a platform for Morais, enriched with both lexicographic and ontological modules.

The lexicographer

- **ANTÓNIO DE MORAIS SILVA** (1757?–1824) was a Brazilian lexicographer born in Rio de Janeiro.

- Morais graduated in Civil and Canon Law from the University of Coimbra in Portugal, but did not embark on a legal career, as he was persecuted by the Inquisition, which, having accused him of heresy, has ordered his arrest.

- He fled to England and devoted himself to studying languages; it was here that he planned the structure of his dictionary.

- Morais moved to Brazil in 1794, where he entered the judiciary and held the position of judge in the Bahia Court of Appeal (from which he soon resigned).

- He subsequently moved to Muribeca, in Pernambuco, where he lived until his death on April 11, 1824.

The dictionary

- These editions are historical Portuguese language dictionaries commonly known by and disseminated under the name of Morais.

- Morais inaugurated modern Portuguese lexicography (1st edition dates of 1789) and was the model for subsequent Portuguese dictionaries throughout the 19th and 20th centuries.

- First edition authorship is attributed to **Rafael Bluteau**, a Portuguese priest and lexicographer, whose Portuguese-Latin Vocabulary (10 vols., 1712-1728) was the basis for the Morais dictionary.

- This work was revised and updated over the years, culminating in the 10th edition.

The editions

1789 (1st ed.)
2 vols.
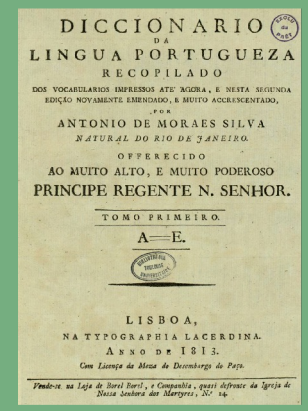
1813 (2nd ed.)
2 vols.
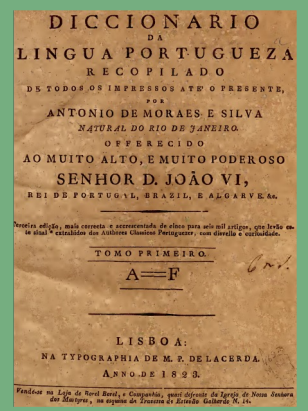
1823 (3rd ed.)
2 vols.
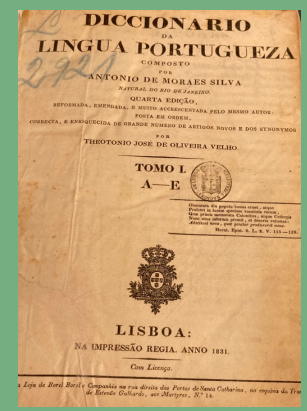
1831 (4th ed.)
2 vols.
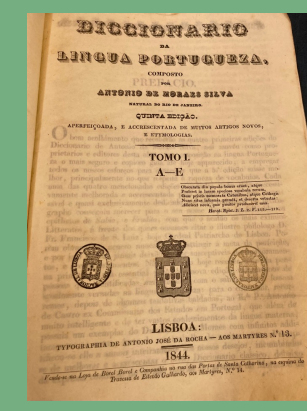
1844 (5th ed.)
2 vols.
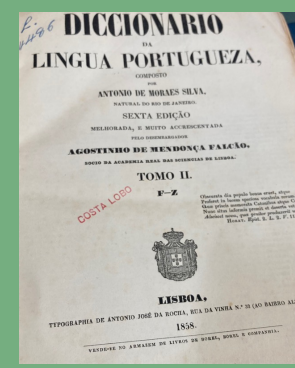
1858 (6th ed.)
2 vols.
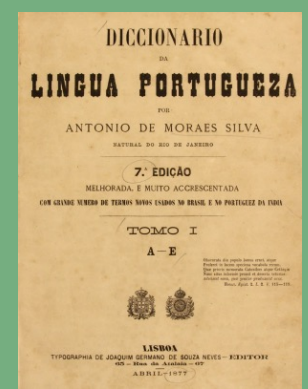
1877-78 (7th ed.)
2 vols.
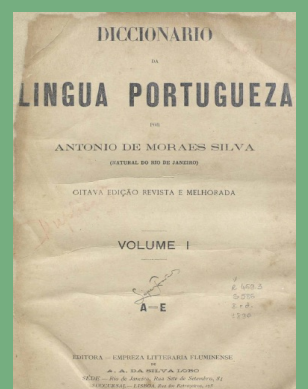
1890-91 (8th ed.)
2 vols.
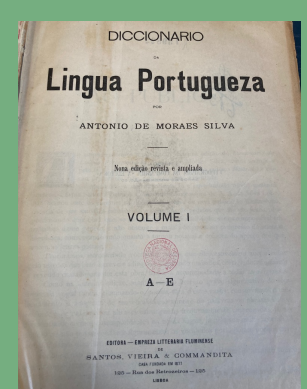
1891 (9th ed.)
2 vols.

1949-59 (10th ed.)
2 vols.

Salgado et al. (2023). Domain labelling in the Morais dictionary

10
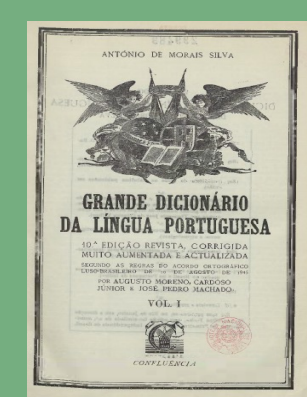
Domain labeling

- We focus on the portion of the lemma list which is made up of **terms**.

- The labels assigned to specialized senses are called **domain labels** (markers that identify the specialized field of knowledge in which a lexical unit is mainly used (Salgado, Costa and Tasovac, 2019).

- A domain (= subject field) is a 'field of knowledge' (ISO 1087:2019).

- De Bessé (2000) distinguishes between (1) domain of knowledge and (2) domain of activities.

- A domain is always an organized set of concepts (Depecker, 2003; Cabré, 1999).

- We believe it would be convenient to establish hierarchical concepts to organize the domains registered in lexicographic resources.

Domain Labeling

- The domain labels were retrieved from the abbreviation lists of its various editions.

- Each table of abbreviations has two distinct columns: one with the abbreviation and the other with the complete domain designations.

- The lists of domains follow an alphabetical ordering, without any concern for the relationships that can be established between those types of labels.

- The list analysis allowed us to develop a typology of usage information in the Morais dictionary: diatechnical, diatextual, diastratic, among many others.

- The classifier 'termo' [term] is used as a textual marker, referring to the domain in which a lexical unit is used.

Fragment of the Morais front matter list (1st edition)

Salgado et al. (2023). Domain labelling in the Morais dictionary

Domain Labeling

35 domain labels

49 domain labels (+ 14)

57 domain labels (+ 8)

59 domain labels (+ 2)

67 domain labels (+ 8)

- number of labels gradually increasing
- constant spelling fluctuation
- change in abbreviations and capitalization
- adjectives > domains

Abbreviation (1–4 ed.)	Full form (1–4 ed.)
Agric.	Agricultura
Anat.	Anatomia, ou Anatómico
Archit.	d'Architectura
Arithm.	Arithmetico
Artelh.	d'Artelharia
Aſtrol.	Aſtrologico
Aſtron.	Aſtronomico
Botan.	Botanico
Braſ.	do Brasão
Chim.	Chimico
Cirurg.	Cirurgico
Eſcult.	Eſcultura
Filoſ.	Filoſofico
Fiſic.	Fiſica
Fortif.	da Fortificação
Geogr.	Geografico
Geometr.	Geometrico
Grammat.	da Grammatica
Jurid.	Juridico
Juriſp.	da Juriſprudencia
Log.	Logico
Manej.	Termo do Manejo dos cavallos
Mathem.	Mathematico
Med.	Medico
Milit.	Militar
Muſ.	Muſico
Naut.	da Nautica
Opt.	Optico
Ortogr.	Ortografico
Perſp.	da Perſpectiva
Pharmac.	Pharmaceutico
Pint.	da Pintura
Rhet.	Rhetorico
Theol.	Theologico
Volat.	Volateria

Abbreviation (5 ed.)	Full form (5 ed.)
Banq.	Banqueiro
Bord.	Bordador
Carp.	Carpinteiro
Comm.	Commercio
Chron. ou Chronol.	Chronologia
Encad.	Encadernador
H. ant.	Historia antiga
h. n. ou nat.	Historia natural
Impr.	Impressor
Merc.	Mercantil
Montar.	Montaria
Myth.	Mythologia, ou Mythologico
Pedr.	Pedreiro
Physiol. ou fisiol.	Physiologia, ou physiologico

Abbreviation (6 ed.)	Full form (6 ed.)
alchym.	Alchymia
did. ou didact.	Didactico
entom.	Entomologia
for. ou forens.	Forense
ichtyol.	Ichtyologia
metall.	Metallurgia
ornit.	Ornithologia
zool.	Zoologia

Abbreviation (7 ed.)	Full form (7 ed.)
ling.	Linguistica
philol.	Philologia

Abbreviation (8 ed.)	Full form (8 ed.)
alg.	Algebra
ant. gr.	Antiguidade Grega
ant. rom.	Antiguidade Romana
epig.	Epigraphia ou Epigraphico
herald.	Heraldica
mar.	Maritimo
mech.	Mechanica
theat.	Theatro

- No modifications!

Domain Labeling

133 domain labels (+ 66)

199 domain labels (+ 106–40=66)

- number of domains increases exponentially
- constant spelling fluctuation
- change in abbreviations
- capitalization

Abbreviation (9 ed.)	Full form (9 ed.)	Abbreviation (9 ed.)	Full form (9 ed.)
acust.	Acustico	hipp.	Hippiatico
alveit.	Alveitario	hortic.	Horticultura
alven.	Alvenaria	hydraul.	Hydraulico
altan.*	Altaneria	indust.	Industria ou Industrial
antig.	Antiguidade	joalh.	Joalheiro
anthrop.	Anthropologico	leg. ou legisl.	Legislativo
archeol.	Archeologico	litter.	Litteratura
ascet.	Ascetico	liturg.	Liturgia, ou Liturgico
banc.	bancario	maçon.	Maçonaria
bell. art.	Bellas artes	meteor.	Meteorologico
biol.	Biologico	metrif.	Metrificação
burocr.	Burocratico	mim.	Mimologia, ou Mimologico
cam. de ferro.	Caminhos de ferro	min. ou miner.	Mineralogia, ou Mineralogico
chirom.	Chiromantico	numis.	Numismatico
choreog.	Choreographico	parl. ou parlam.	Parlamentar
chorog.	Chorographico	pathol.	Pathologico
conchyl.	Conchyliologico	phon.	Phonologia
cul. ou culin.	Culinario	photog.	Photographico
diplom.	Diplomatico	pyrot.	Pyrotecnico
dir.	Direito	reloj.	Relojoeiro
dog.	Dogmático	sapat.	Sapateiro
dour.	Dourador	serralh.	Serralheiro
eccles.	Eclesiástico	tact.	Tactica
econ. pol.	Economia politica	tan.	Tanoeiro
eng.	Engenharia	taurom.	Tauromachico
equit.	Equitação	techn.	Techonologia
eschol.	Eschola, ou Escholastico	terat.	Teratologico
esgr.	Esgrima	therap.	Therapeutico
fin. ou financ.	Financeiro	typ. ou typog.	Typographico
fund.	Fundidor	versif.	Versificação
geod.	Geodesico	veter. ou veterin.	Veterinario
geol.	Geologia	vinic.	Vinicultura
grav.	Gravura, ou Gravador	zootechn.	Zootechnico

Abbreviation (10 ed.)	Full form (10 ed.)	Abbreviation (10 ed.)	Full form (10 ed.)	Abbreviation (10 ed.)	Full form (10 ed.)
Açoug.	Açougueiro	Dec.	Decorativo (va)	Ocean.	Oceanografia
Adm.	Administração	Des.	Desenho	Oneol.	Oneologia
Aeron.	Aeronáutica	Desp.	Desportos	Paleog.	Paleografia
Aerost.	Aerostática	Dinâm.	Dinâmica	Paleont.	Paleontologia
Agron.	Agronomia	Dir. rom.	Direito romano	Pedag.	Pedagogia
Alfaiat.	Alfaiataria	Electr.	Electricidade	Pesc.	Pescaria
Alpin.	Alpinismo	Embr.	Embriologia	Petr.	Petrologia
Alquil.	Alquilaria	Enol.	Enologia	Pisc.	Piscicultura
Apic.	Apicultura	Espir.	Espiritismo	Polít.	Política
Arbor.	Arboricultura	Estat.	Estatística	Pré-hist.	Pré-história
Arquit. hidr.	Arquitectura hidráulica	Etnogr.	Etnografia	Pros.	Prosódia
Arquit. mil.	Arquitectura militar	Falc.	Falcoaria	Psicol.	Psicologia
Arquit. nav.	Arquitectura naval	Feud.	Feudalismo	Psiq.	Psiquiatria
Arm.	Armaria	Folc.	Folclore	Radiol.	Radiologia
Arte mil.	Arte militar	Fonét.	Fonética	Radiot.	Radiotecnia
Autom.	Automóvel ou automobilismo	Geneal.	Genealogia	Rel.	Religião
Av.	Aviação	Gin.	Ginástica	Rev.	Revisão
Bact.	Bacteriologia	Ginec.	Ginecologia	Sal.	Salinas
Balíst.	Balística	Herm.	Hermetismo	Semân.	Semântica
Bibl.	Bíblia, biblionismo e bibliónimo	Hidrogr.	Hidrografia	Seric.	Sericicultura
Cabal.	Cabalística	Hig.	Higiene	Sid.	Siderotecnia
Cal.	Caligrafia	Hist.	História	Silv.	Silvicultura
Cartog.	Cartografia	Hist. Rel.	História Religiosa	Sint.	Sintaxe
Catól.	Católico	Histol.	Histologia	Sociol.	Sociologia
Cav.	Cavalaria	Homeop.	Homeopatia	Teleg.	Telegrafia
Cerâm.	Cerâmica	Jard.	Jardinagem	Territ.	Território
Chancel.	Chancelaria	Lot.	Lotaria	Tint.	Tinturaria
Chapel.	Chapelaria	Magn.	Magnetismo	Toc.	Tocologia
Cineg.	Cinegético ou Cinegética	Marc.	Marcenaria	Top.	Topografia
Constr.	Construção	Marn.	Marnoto	Toxic.	Toxicologia
Cosmog.	Cosmografia	Metaps.	Metapsíquica	Trigon.	Trigonometria
Cosmol.	Cosmologia	Micr.	Microscopia	T. S. F.	Telegrafia sem fios
Crist.	Cristalografia	Mor.	Moral	Veloc.	Velocipedia
Curt.	Curtume	Nav.	Naval	Ven.	Venatório
Cutel.	Cutelaria	Obst.	Obstetrícia	Vin.	Vinificação
				Vit.	Viticultura

antig.	Antiguidade
ant. gr.	Antiguidade grega
ant. rom.	Antiguidade romana
ascet.	Ascetico
banc.	bancario
banq.	banqueiro
bell. art.	Bellas artes
bord.	Bordador
braz.	Brazão
cam. de ferro.	Caminhos de ferro
chirom.	Chiromantico
conchyl.	Conchyliologico
dog.	Dogmático
dour.	Dourador
encad.	Encadernador
esgr.	Esgrima
for. ou forens.	Forense
fort. ou fortif.	Fortificação
fund.	Fundidor
grav.	Gravura, ou Gravador
hipp.	Hippiatico
h. ant.	Historia antiga
hortic.	Horticultura
indust.	Industria ou Industrial
joalh.	Joalheiro
Jurispr.	Jurisprudencia
leg. ou legisl.	Legislativo
manej.	Manejo dos cavallos
merc.	Mercantil
metrif.	Metrificação
mim.	Mimologia, ou Mimologico
myth.	Mythologia, ou Mythologico
pedr.	Pedreiro
persp.	Perspectiva
phon.	Phonologia
reloj.	Relojoeiro
sapat.	Sapateiro
tact.	Tactica
tan.	Tanoeiro
volat.	Volateria

Salgado et al. (2023). Domain labelling in the Morais dictionary

Domain Labeling

(1, 2, 3, 4 ed.)

(6 ed.)

Abbreviation (1-2 ed.)	Full form (1-2 ed.)	Abbreviation (3-4 ed.)	Full form (3-4 ed.)	Abbreviation (5-9 ed.)
Braſ.	do Brasão	Bras.	Brasão	Braz.

Abbreviation (1 ed.)	Full form (1 ed.)	Abbreviation (2 ed.)	Full form (2 ed.)	Abbreviation (3-4 ed.)	Full form (3-4 ed.)		
Filoſ.	Filoſofico	Filos.	Filosofia, ou Filosofico	Filos.	Filosofia, ou Filosofico		
Abbreviation (5 ed.)	**Full form (5 ed.)**	**Abbreviation (6-7-8 ed.)**	**Full form (6-7-8 ed.)**	**Abbreviation (9 ed.)**	**Full form (9 ed.)**	**Abbreviation (10 ed.)**	**Full form (10 ed.)**
philos. ou filos.	Philosofia, ou philosofico	philos.	Philosofia, ou Philosofico	phil.	Philosofico	Filos.	Filosofia

Full forms (1–4 ed.)

Knowledge organization

- We look for a possible organizational structure to help identify generic domains and subdomains, trying to establish a hierarchical structure.

- In organizing domains, we propose three possible levels: superdomain, domain, and subdomain.

- This hierarchy allows to organize an increasing amount of terminological data and provides greater control over specialized content.

Salgado et al. (2023). Domain labelling in the Morais dictionary

Salgado (2021). Terminological Methods in Lexicography: Conceptualising, Organising and Encoding Terms in General Language Dictionaries
Costa et al. (2023). A crossroad between lexicography and terminology work: knowledge organisation and domain labelling

Knowledge organization

Flat domain labels

Knowledge organization

Hierarchical domain labels
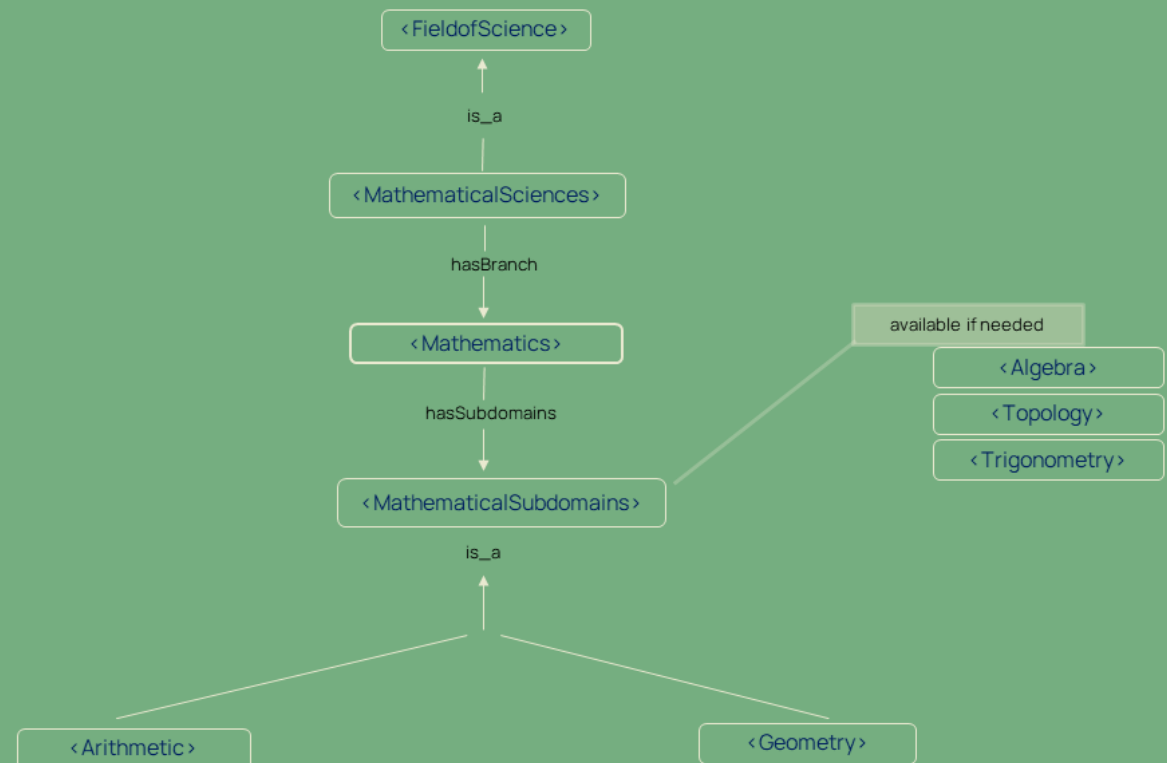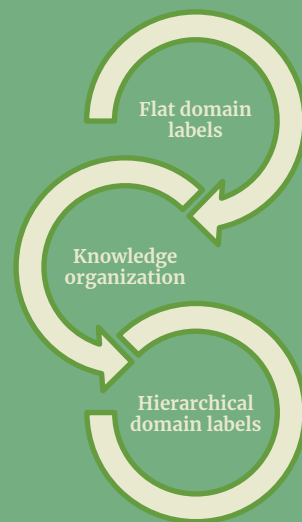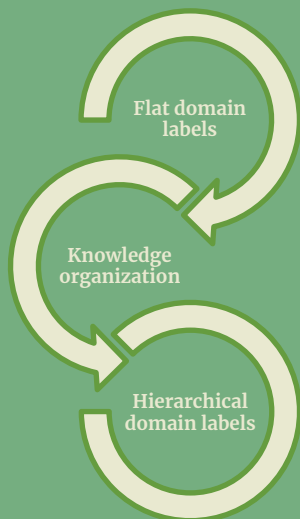
```xml
<taxonomy xml:id="domains">
    <category xml:id="domain.mathematical_sciences"
    valueDatcat="http://www.semanticweb.org/OntoDomLab-Math#MathematicalSciences
http://vocabs.rossio.fcsh.unl.pt/morais_domains/0036">
    <catDesc xml:lang="en">
        <term>Mathematical Sciences</term>
        <!-- Source of gloss: https://dbpedia.org/page/Mathematical_sciences-->
        <gloss>Group of areas of study that includes, in addition to mathematics, those
            academic disciplines that are primarily mathematical in nature but may not
            be universally considered subfields of mathematics proper.</gloss>
    </catDesc>
    <catDesc xml:lang="pt">
        <term>Ciências Matemáticas</term>
    </catDesc>
    <category xml:id="domain.mathematics"
        valueDatcat="http://www.semanticweb.org/OntoDomLab-Math#Mathematics
http://vocabs.rossio.fcsh.unl.pt/morais_domains/0024">
        <catDesc xml:lang="en">
            <term>Mathematics</term>
        </catDesc>
        <catDesc xml:lang="pt">
            <term>Matemática</term>
        </catDesc>
        <category xml:id="domain.arithmetic"
            valueDatcat="http://www.semanticweb.org/OntoDomLab-Math#Arithmetic
http://vocabs.rossio.fcsh.unl.pt/morais_domains/0003">
            <catDesc xml:lang="en">
                <term>Arithmetic</term>
            </catDesc>
            <catDesc xml:lang="pt">
                <term>Aritmética</term>
            </catDesc>
        </category>
        <category xml:id="domain.geometry"
            valueDatcat="http://www.semanticweb.org/OntoDomLab-Math#Geometry
http://vocabs.rossio.fcsh.unl.pt/morais_domains/0018">
            <catDesc xml:lang="en">
                <term>Geometry</term>
            </catDesc>
            <catDesc xml:lang="pt">
                <term>Geometria</term>
            </catDesc>
        </category>
    </category>
    </category>
</taxonomy>
```

Salgado et al. (2023). Domain labelling in the Morais dictionary

Salgado (2021). Terminological Methods in Lexicography: Conceptualising, Organising and Encoding Terms in General Language Dictionaries
Costa et al. (2023). A crossroad between lexicography and terminology work: knowledge organisation and domain labelling

Knowledge organization

- To facilitate the analysis and to focus on interoperability issues, we generate a metalabel, i.e., a tag that identifies the English equivalent of the corresponding domain.

- The survey of all domain labels allows us to determine the number of domains represented (239).

- We found (1) domains of knowledge (2) domains of activities (professional and leisure) (3) domains of jobs.

Domain labels in Morais dictionary

METALABEL	DOMAIN	METALABEL	DOMAIN	METALABEL	DOMAIN	METALABEL	DOMAIN
Acoustics	Acústica	Cynegetics	Cinegético ou Cinegética	Law	Direito	Press	Imprensa
Administration	Administração	Decorative	Decorativo (va)	Legislation	Legislação	Proofreading	Revisão
Aeronautics	Aeronáutica	Diplomacy	Diplomática, diplomacia	Legislation	Legislativo	Prosody	Prosódia
Aerostatics	Aerostática	Dogma	Dogmático	Linguistics	Linguística	Psychiatry	Psiquiatria
Agriculture	Agricultura	Drawing	Desenho	Literature	Literatura	Psychology	Psicologia
Agronomy	Agronomia	Dye	Tinturaria	Liturgy	Liturgia	Pyrotechnics	Pirotecnia
Alchemy	Alquimia	Dynamics	Dinâmica	Logic	Lógica	Radiology	Radiologia
Algebra	Álgebra	Ecclesiastical	Eclesiástico	Lottery	Lotaria	Radiotechnics	Radiotecnia
Alpinism	Alpinismo	Electricity	Electricidade	Magnetism	Magnetismo	Railway	Caminhos de ferro
Anatomy	Anatomia	Embroiderer	Bordador	Maritime	Marítimo	Religion	Religião
Ancient history	Historia antiga	Embryology	Embriologia	Masonry	Alvenaria	Religious History	História Religiosa
Animal healing	Alveitar, alveitaria	Engeneering	Engenharia	Masonry	Maçonaria	Rhetoric	Retórica
Anthropology	Antropologia	Entomatology	Entomatologia	Mathematics	Matemática	Riding	Equitação
Antiquity	Antiguidade	Epigraphy	Epigrafia	Mechanics	Mecânica	Roman Antiquity	Antiguidade romana
Apiculture	Apicultura	Equestrianism	Manejo dos cavallos	Medicine	Medicina	Roman law	Direito romano
Arboriculture	Arboricultura	Ethnography	Etnografia	Mercantile	Mercantil	Saltern worker	Marnoto
Archaeology	Arqueologia	Falconry	Falcoaria	Metallurgy	Metalurgia	Saltworks	Salinas
Architecture	Arquitectura	Falconry	Volateria	Metapsychics	Metapsíquica	Scholasticism	Escolástica
Arithmetic	Aritmética	Fencing	Esgrima	Meteorology	Meteorologia	Sculpture	Escultura
Armoury	Armaria	Feudalism	Feudalismo	Metrification	Metrificação	Seamanship	Náutica
Art of teaching	Didáctica	Finance	Finanças	Microscopy	Microscopia	Semantics	Semântica
Artillery	Artilharia	Fine Arts	Bellas artes	Military	Militar	Sericulture	Sericicultura
Asceticism	Asceto	Fishing	Pescaria	Military architecture	Arquitectura militar	Shoe repairer	Sapateiro
Astrology	Astrologia	Folklore	Folclore	Military art	Arte militar	Siderotechny	Siderotecnia
Astronomy	Astronomia, astrónimo	Forensic	Forense	Mimicking	Mimologia, ou Mimologico	Silviculture	Silvicultura
Aviation	Aviação	Fortification	Fortificação	Mineralogy	Mineralogia	Slaughterer	Açougueiro
Bacteriology	Bacteriologia	Gardening	Jardinagem	Moral	Moral	Sociology	Sociologia
Ballistics	Balística	Genealogy	Genealogia	Motor racing	Automóvel ou automobilismo	Spiritualism	Espiritismo
Banker	banqueiro	Geodesy	Geodesia	Music	Música	Sport	Desportos
Banking	bancario	Geography	Geografia	Mythology	Mythologia, ou Mythologico	Statistics	Estatística
Bible	Bíblia, biblionismo e bibliónimo	Geology	Geologia	Natural history	História Natural	Stonemason	Pedreiro
Biology	Biologia	Geometry	Geometria	Naval	Naval	Surgery	Cirurgia
Blacksmith's art	Serralharia	Gilder	Dourador	Naval architecture	Arquitectura naval	Syntax	Sintaxe
Blazon	Brazão	Grammar	Gramática	Numismatics	Numismática	Tactics	Tactica
Bookbinder	Encadernador	Gravure	Gravura, ou Gravador	Obstetrics	Obstetrícia	Tailor's	Alfaiataria
Botany	Botânica	Greek Antiquity	Antiguidade grega	Oceanography	Oceanografia	Tanning	Curtume
Bullfighting	Tauromaquia	Gymnastics	Ginástica	Oenology	Enologia	Technology	Tecnologia
Bureaucracy	Burocracia	Gynaecology	Ginecologia	Oneology	Oneologia	Telegraphy	Telegrafia
Calligraphy	Caligrafia	Hatmaking	Chapelaria	Optics	Óptica	Teratology	Teratologia
Carpentry	Carpintaria	Hawking	Altanaria	Ornithology	Ornitologia	Territory	Território
Cartography	Cartografia	Heraldry	Heráldica	Orthography	Ortografia	Theatre	Teatro
Caster	Fundidor	Hermetism	Hermetismo	Painting	Pintura	Theology	Teologia
Catholic	Católico	Hippiatrics	Hipiátrico	Palaeography	Paleografia	Therapeutics	Terapêutica
Cavalry	Cavalaria	Hiring place	Alquilaria	Palaeontology	Paleontologia	Tocology	Tocologia
Ceramics	Cerâmica	Histology	Histologia	Parliament	Parlamento ou parlamentarismo	Topography	Topografia
Chancellery	Chancelaria	History	História	Pathology	Patologia	Toxicology	Toxicologia
Chemistry	Química	Homeopathy	Homeopatia	Pedagogy	Pedagogia	Trade	Comércio
Chiromancy	Chiromantico	Horticulture	Horticultura	Perspective	Perspectiva	Trigonometry	Trigonometria
Choreography	Coreografia	Hunting	Montaria	Petrology	Petrologia	Typography	Tipografia
Chorography	Corografia	Hunting	Venatório	Pharmacy	Farmácia	Velocipedia	Velocipedia
Chronology	Cronologia	Hydraulic architecture	Arquitectura hidráulica	Philology	Filologia	Versification	Versificação
Conchology	Conchyliologico	Hydraulics	Hidráulica	Philosophy	Filosofia	Veterinary	Veterinária
Construction	Construção	Hydrography	Hidrografia	Phonetics	Fonética	Viniculture	Vinicultura
Cooper	Tanoeiro	Hygiene	Higiene	Phonology	Fonologia	Vinification	Vinificação
Cosmography	Cosmografia	Ichthyology	Ictiologia	Photography	Fotografia	Viticulture	Viticultura
Cosmology	Cosmologia	Industry	Industria ou Industrial	Physics	Física	Watchmaker	Relojoeiro
Crystallography	Cristalografia	Jeweller	Joalheiro	Physiology	Fisiologia	Wireless telegraphy	Telegrafia sem fios
Cuisine	Culinária	Jurisprudence	Jurídico e Jurisprudência	Pisciculture	Piscicultura	Woodwork	Marcenaria
Cutlery	Cutelaria	Kabbalistic	Cabalística	Political economy	Economia Política	Zoology	Zoologia, ou Zoologico
				Politics	Política	Zootechny	Zootecnia
				Prehistory	Pré-história		

Knowledge organization

MorDigital

MATHEMATICA , ſ. f. a ſciencia , que enſina a conhecer as grandezas de toda ſorte , ſuas razões , relações , e proporções : *Mathematica miſta* (oppõe-ſe ás *puras*) a que enſina a applicar os principios de calculo , e geometria aos corpos.
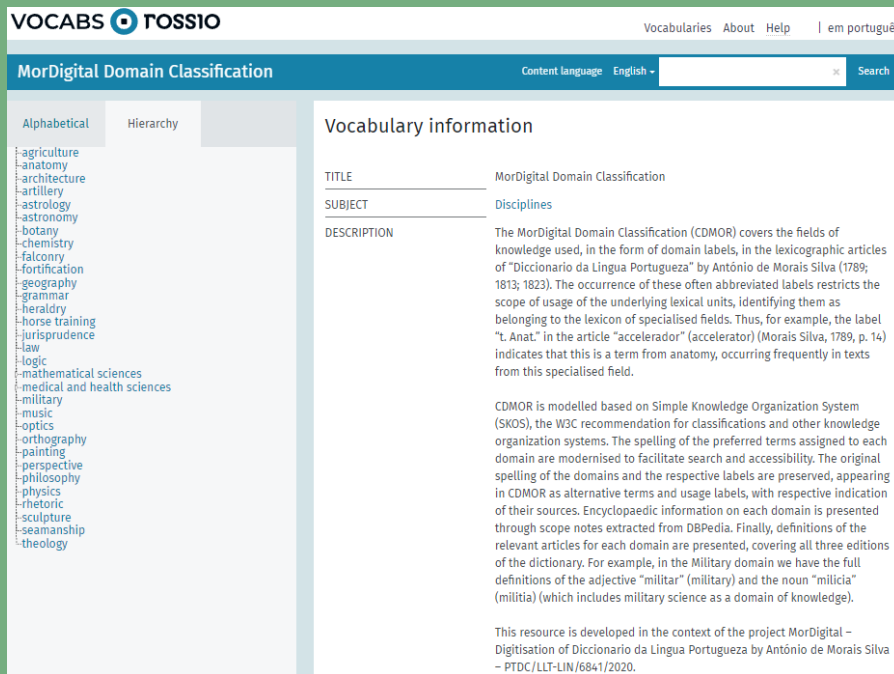
```
ORDENADA , ſ. f. Mathem. linha recta tirada perpendicularmente do ponto da curva a ſeu eixo.
```

```
<entry xmlns="http://www.tei-c.org/ns/1.0" xml:id="MOR1.DLP.ORDENADA" type="mainEntry" xml:lang="pt">
    <form type="lemma">
        <orth>ORDENADA</orth>
    </form>
    <metamark function="lemmaDelimiter">,</metamark>
    <gramGrp>
        <gram type="pos" norm="NOUN">ſ.</gram>
        <gram type="gen">f.</gram>
    </gramGrp>
    <sense xml:id="MOR1.DLP.ORDENADA.s.1">
        <usg type="domain" corresp="#domain.mathematics">Mathem.</usg>
        <def>linha recta tirada perpendicularmente do ponto da curva a ſeu eixo</def>
    </sense>
    <pc>.</pc>
</entry>
```

TRIGONOMETRIA , ſ. f. parte da Mathematica , que enſina a reſolver os triangulos planos , e esfericos.

```
<entry xmlns="http://www.tei-c.org/ns/1.0" xml:id="MOR1.DLP.TRIGONOMETRIA"
type="mainEntry" xml:lang="pt">
    <form type="lemma">
        <orth>TRIGONOMETRIA</orth>
    </form>
    <metamark function="lemmaDelimiter">,</metamark>
    <gramGrp>
        <gram type="pos" norm="NOUN">ſ.</gram>
        <gram type="gen">f.</gram>
    </gramGrp>
    <sense xml:id="MOR1.DLP.TRIGONOMETRIA.s.1">
        <usg type="domain" corresp="#domain.mathematics" resp="#Salgado"/>
        <def>parte da Mathematica , que enſina a reſolver os triangulos planos , e
esfericos</def>
    </sense>
    <pc>.</pc>
</entry>
```

```
APOTEMA , ſ. m. Matemat. raio recto v. g. ,, a apotema de hum poligono he a recta perpendicularmente tirada do centro ao lado do poligono.
```

```
<entry xmlns="http://www.tei-c.org/ns/1.0" xml:id="MOR1.DLP.APOTEMA" type="mainEntry" xml:lang="pt">
    <form type="lemma">
        <orth>APOTEMA</orth>
    </form>
    <metamark function="lemmaDelimiter">,</metamark>
    <gramGrp>
        <gram type="pos" norm="NOUN">ſ.</gram>
        <gram type="gen">m.</gram>
    </gramGrp>
    <sense xml:id="MOR1.DLP.APOTEMA.s.1">
        <usg type="domain" corresp="#domain.mathematics">Matemat.</usg>
        <def>raio recto</def>
        <lbl expand="verbi gratia" xml:lang="la"><hi>v. g.</hi></lbl>
        <metamark function="exampleDelimiter">,,</metamark>
        <cit type="example" xml:lang="pt">
            <quote>a apotema de hum poligono he a recta perpendicularmente tirada do centro ao
                lado do poligono</quote>
        </cit>
    </sense>
    <pc>.</pc>
</entry>
```

Salgado et al. (2023). Domain labelling in the Morais dictionary

Tasovac et al. (2018). TEI Lex-0: A baseline encoding for lexicographic data. Version 0.9.2. DARIAH Working Group on Lexical Resources.
https://dariah-eric.github.io/lexicalresources/pages/TEILex0/TEILex0.html

Knowledge organization

https://vocabs.rossio.fcsh.unl.pt/pub/morais_domains/en/

https://github.com/GuidaRamos/OntoDomLab-Math

Salgado et al. (2023). Domain labelling in the Morais dictionary
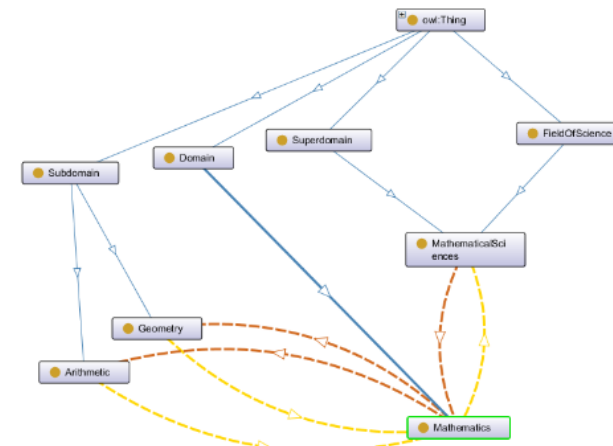
Almeida et al. (2022). MorDigital Domain Classification (CDMOR). http://vocabs.rossio.fcsh.unl.pt/morais_domains/
Ramos et al. (2023). OntoDomLab-Math: domain label ontology focused on Mathematics and its subdomains. https://github.com/GuidaRamos/OntoDomLab-Math

MorDigital

Final remarks

- Domain organization is crucial to improve the labeling system in dictionaries.

- The hierarchical structure allows to organize an increasing amount of terminological data and provides greater control over specialized content.

- Combining an onomasiological with a semasiological approach and its respective encoding can be applied to the future search engine.

- TEI Lex-0 enables the XML encoding of the Morais dictionary in an interoperable way, providing the basis for visualizing, querying, and mining dictionary data in the MorDigital platform.

- The methodologies and resources developed for MorDigital should be reusable in other projects involving the retrodigitization of legacy dictionaries.

- By observing various lexicographic resources, we believe we are taking an essential step toward a possible homogenization of the representation of lexicographic data.

Acknowledgements

MorDigital

MorDigital

Thank you!

Ana Salgado: anasalgado@fcsh.unl.pt

Rute Costa: rute.costa@fcsh.unl.pt

Toma Tasovac: ttasovac@humanistika.org

Anas Fahad Khan: fahad.khan@ilc.cnr.it

Bruno Almeida: brunoalmeida@fcsh.unl.pt

Margarida Ramos: mvramos@fcsh.unl.pt

Sara Carvalho: sara.carvalho@ua.pt

Raquel Silva: raq.silva@fcsh.unl.pt

Mohamed Khemakhem: medkhemakhemfsegs@gmail.com

Laurent Romary: laurent.romary@inria.fr

CLUNL
CENTRO DE LINGUÍSTICA DA UNIVERSIDADE NOVA DE LISBOA

NOVAFCSH
FACULDADE DE CIÊNCIAS SOCIAIS E HUMANAS
UNIVERSIDADE NOVA DE LISBOA

FCT Fundação
para a Ciência
e a Tecnologia