Ciclo de Encontros: 4 x PLN

Humanidades Digitais e PLN

Renata Vieira, Fernanda Olival, Helena Cameron

28 de outubro 2024

Fátima Farrica
Maria José Bocorny Finatto
Ana Paula Banza
Ana Sofia Ribeiro
Cassia Trojahn

Sumário

imagem

Introdução

Introdução

Na área de HD, em fontes textuais, encontramos uma grande variação

- nos **períodos históricos** das fontes,
- no seu suporte (**manuscritos em papel, impressos, fotografados**, etc),
- no seu estágio de digitalização, que pode variar entre i**magens digitais, textos em PDF e textos digitalizados** em outros formatos.

Todas essas variações adicionam esforços extras de processamento.

Introdução

O capítulo discute os requisitos de **preparação e organização das fontes,** objetos de análise que depois de transcritas e **digitalizadas**, podem ser submetidas a **processamentos mais avançados**.

O objetivo é mostrar não apenas como o **PLN é útil e relevante nesse domínio**, mas também como a área de **HD é rica em despertar novas questões** para o desenvolvimento do PLN.

Preparação das Fontes

Digitalização

Normalização

Metadados

Anotação

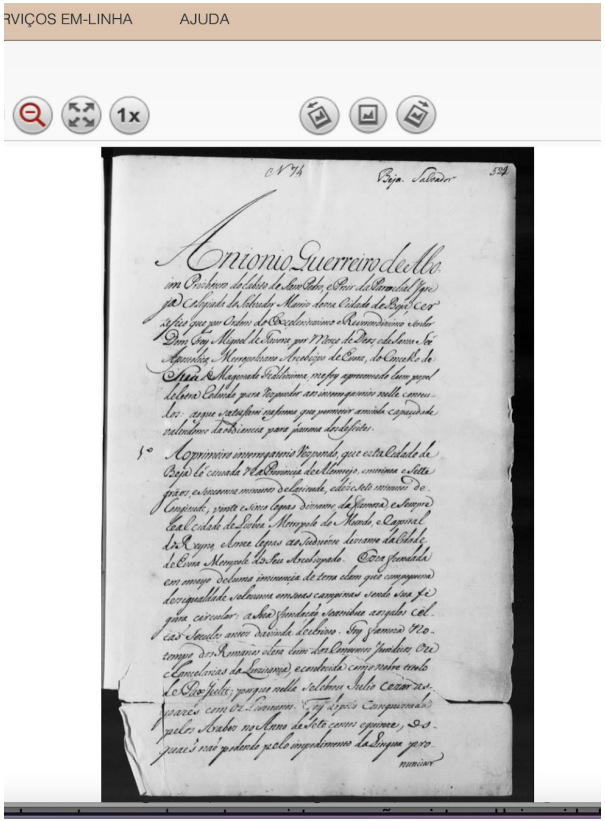
Digitalização

Muitos projetos na área de HD necessitam lidar com a digitalização de **manuscritos** originais.

A área de paleografia, em particular, lida com a leitura e transcrição de versões manuscritas para um suporte atual.

Paleografia manual x automática (ex. Transkribus)

Digitalizar impressos (ex. eScriptorium, vFlat scan)



Metadados

Uma vez digitalizada a fonte ou corpus de estudo, é necessário pensar na **organização dos seus metadados.**

O material digital deve conter a informação sobre a qual acervo pertence, identificar unicamente cada arquivo e, quando pertinente, associar autoria, data e outros elementos pertinentes.

Os metadados podem descrever a estrutura do documento, identificar volumes, capítulos, páginas, cabeçalhos, notas de rodapé, numeração de páginas ou comentários adicionados aos originais.

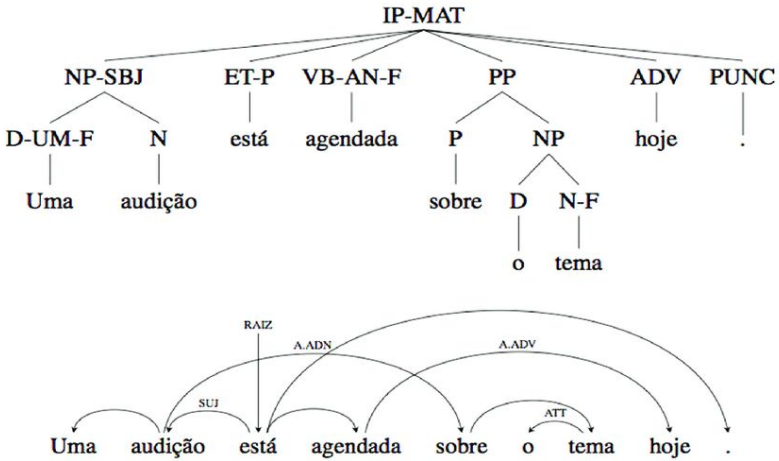
Normalização

Em HD, as fontes textuais, quer manuscritas, quer impressas, sendo de uma época anterior ao estágio atual da língua, apresentam **variações ortográficas ou morfossintáticas**, não só em relação ao padrão atual como também dentro da mesma época.

Anotação

A anotação textual ocorre em diferentes níveis, **sintático, semântico lexical, e discursivo** para identificar elementos de interesse e embasar diferentes estudos.

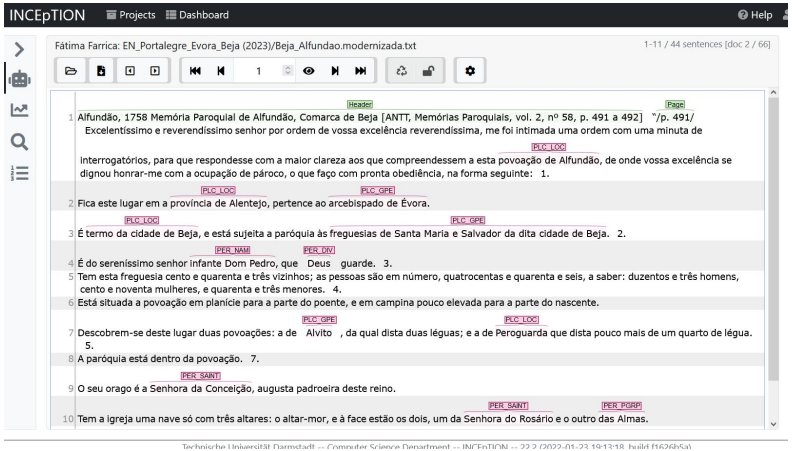
Geralmente a anotação sintática permite um tipo de análise diferenciada da anotação semântica.



Anotação

A anotação de **entidades** é frequentemente adotada, com utilidade para estudos de literatura, história, geografia...

Conforme a complexidade do fenômeno e a respectiva disponibilidade de recursos, a anotação pode ser manual ou automática.

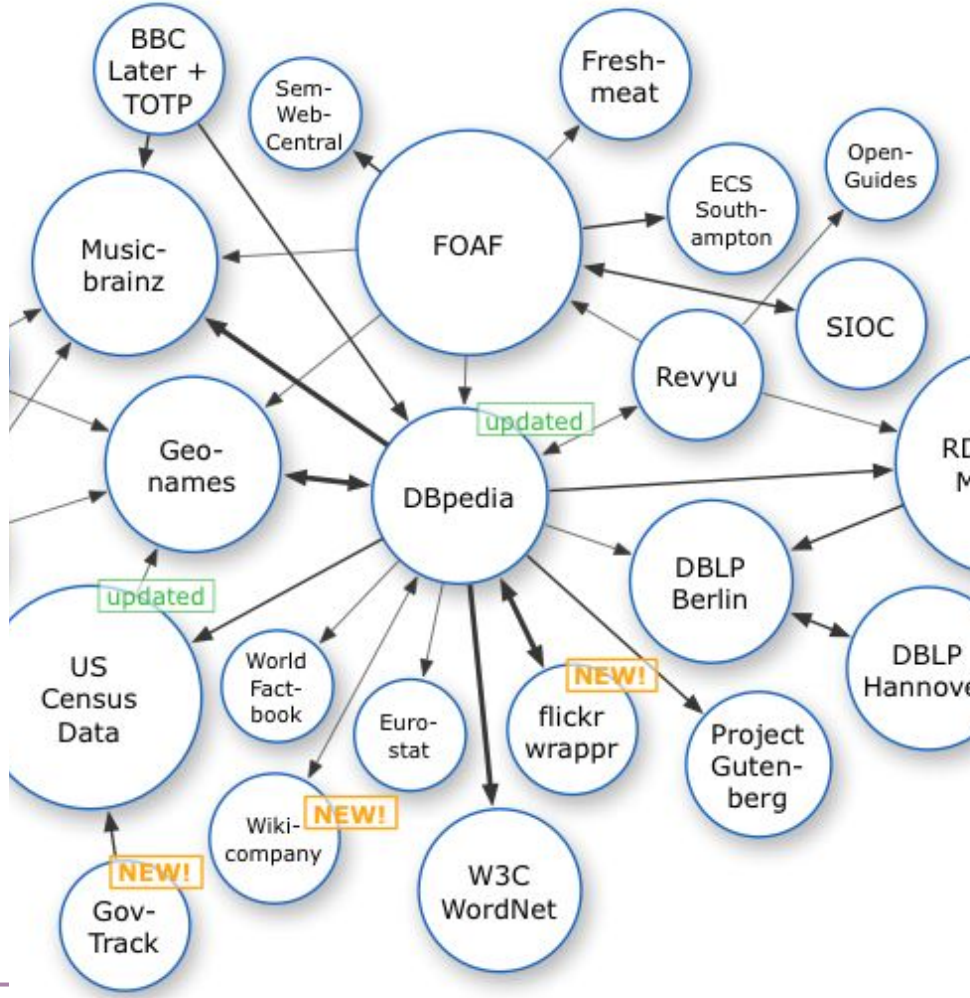


Transformação de Texto em Dados

Extração de Informação

Representação de conhecimento e Ontologias

Dados Ligados e Dados FAIR



Extração de Informação

O PLN atua como área responsável pela transformação de **textos em dados**.

Em textos das áreas das humanidades já digitalizados e normalizados, podem ser aplicadas ferramentas de PLN já desenvolvidas para diferentes tarefas:

- reconhecimento de **entidades nomeadas**,
- extração de **eventos**,
- resolução de **correferências**, e
- sistemas de **respostas a perguntas**

Amplia a **capacidade de análise,** pois pode-se organizar essa informação de forma ágil e rápida em estruturas bem definidas.

Extração de Informação

Código	Nome	Ligação ao SO	Ofício	Morada_prc
	Diogo Pereira de Sampaio	Processo incompleto		Beira
14819	Domingos Fernandes	Familiar SO	Mercador de panos	Beira
	Domingos João	Familiar SO	Mercador	Beira
	Estêvão Fernandes	Familiar SO	Barbeiro	Estremadura
20437	Estêvão Lopes	Familiar SO	Tanoeiro	Estremadura
	Estêvão Pina	Familiar SO	Boticário	Estremadura
	Filipe Lourenço, Padre	Notário SO	Escrivão do Comissário SO	Beira
	Francisco de Araújo	Oficial SO	Despenseiro SO	
	Francisco Gomes de Melo, Padre	Notário SO	Notário SO junto do comiss	Beira
	Francisco Gonçalves	Familiar SO	Alfaiate	Estremadura

Representação do Conhecimento e Ontologias

Ontologias têm sido usadas em diversos domínios do conhecimento em diferentes tarefas:

- representar conhecimento de **domínios** complexos;
- organizar e anotar grandes quantidades de dados (**anotação semântica**);
- integrar dados de diversas fontes (**integração** semântica);
- dar ancoragem para o **raciocínio automático**;
- suportar maior **eficiência semântica** em buscas.

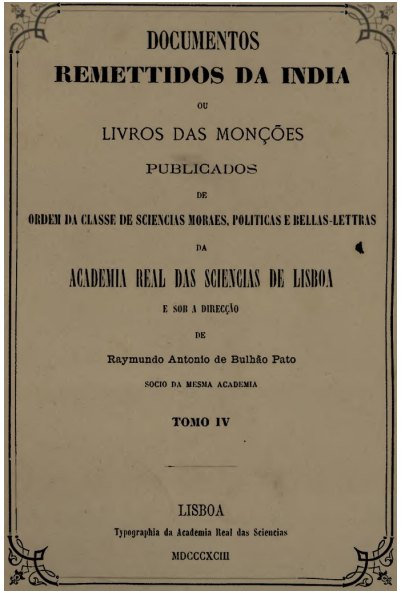
Dados Ligados e Dados FAIR

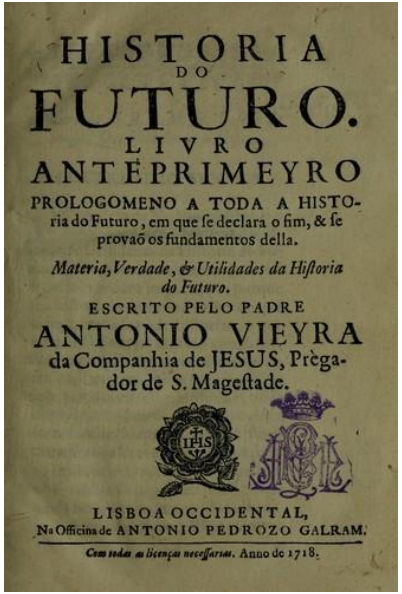
Findable, Accessible, Interoperable, Reusable

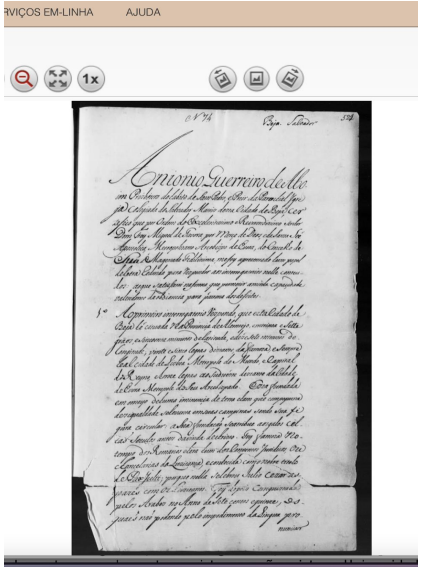
- (F) atribuição de i**dentificadores exclusivos** e persistentes a conjuntos de dados, descrevendo-os com metadados ricos que permitem sua indexação e descoberta;
- (A) uso de **protocolos abertos e de padrões** para acesso a conjuntos de dados;
- (I) uso de linguagens formais e de **vocabulários padronizados** FAIR para representar (meta)dados;
- (R) uso de metadados ricos sobre **licença de uso, proveniência e qualidade** de dados.

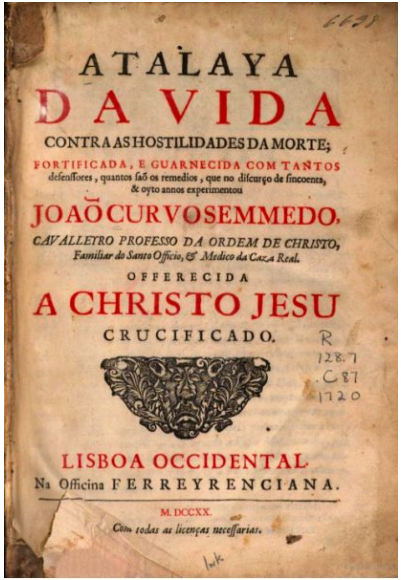
PLN, HD e Língua Portuguesa:

Projetos

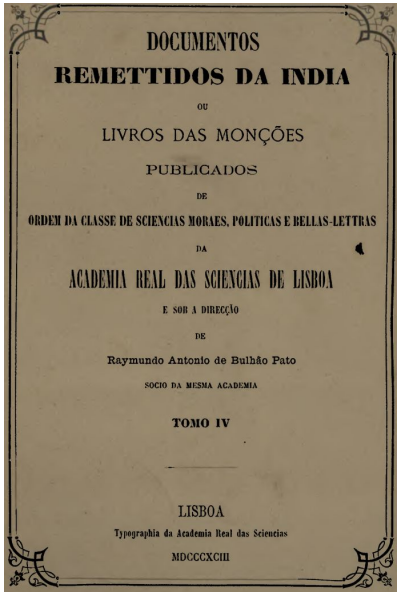

DOCUMENTOS
REMETTIDOS DA INDIA
OU
LIVROS DAS MONÇÕES
PUBLICADOS
DE
ORDEM DA CLASSE DE SCIENCIAS MORAES, POLITICAS E BELLAS-LETTRAS
DA
ACADEMIA REAL DAS SCIENCIAS DE LISBOA
E SOB A DIRECÇÃO
DE
Raymundo Antonio de Bulhão Pato
SOCIO DA MESMA ACADEMIA
TOMO IV
LISBOA
Typographia da Academia Real das Sciencias
MDCCCXCIII


HISTORIA
DO
FUTURO.
LIVRO
ANTEPRIMEYRO
PROLOGOMENO A TODA A HISTORIA do Futuro, em que se declara o fim, & se provaõ os fundamentos della.
Materia, Verdade, & Utilidades da Historia do Futuro.
ESCRITO PELO PADRE
ANTONIO VIEYRA
da Companhia de JESUS, Prègador de S. Magestade.
LISBOA OCCIDENTAL,
Na Officina de ANTONIO PEDROZO GALRAM.
Com todas as licenças necessarias. Anno de 1718.



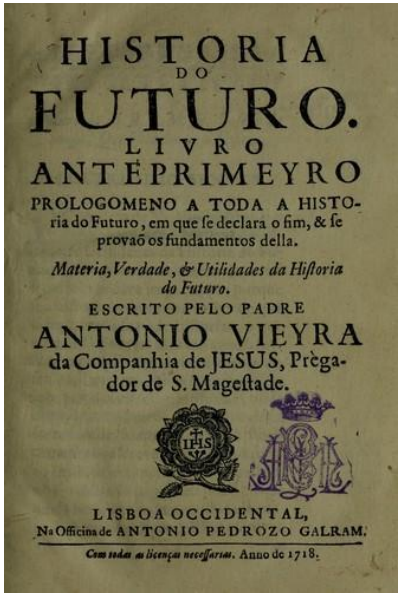

ATALAYA
DA VIDA
CONTRA AS HOSTILIDADES DA MORTE;
FORTIFICADA, E GUARNECIDA COM TANTOS defensores, quantos saõ os remedios, que no discurço de sincoenta, & oyto annos experimentou
JOAÕ CURVO SEMMEDO,
CAVALLEIRO PROFESSO DA ORDEM DE CHRISTO, Familiar do Santo Officio, & Medico da Caza Real.
OFFERECIDA
A CHRISTO JESU
CRUCIFICADO.
LISBOA OCCIDENTAL
Na Officina FERREYRENCIANA.
M. DCCXX.
Com todas as licenças necessarias.

Nossas fontes


DOCUMENTOS
REMETTIDOS DA INDIA
OU
LIVROS DAS MONÇÕES
PUBLICADOS
DE
ORDEM DA CLASSE DE SCIENCIAS MORAES, POLITICAS E BELLAS-LETTRAS
DA
ACADEMIA REAL DAS SCIENCIAS DE LISBOA
E SOB A DIRECÇÃO
DE
Raymundo Antonio de Bulhão Pato
SOCIO DA MESMA ACADEMIA
TOMO IV
LISBOA
Typographia da Academia Real das Sciencias
MDCCCXCIII

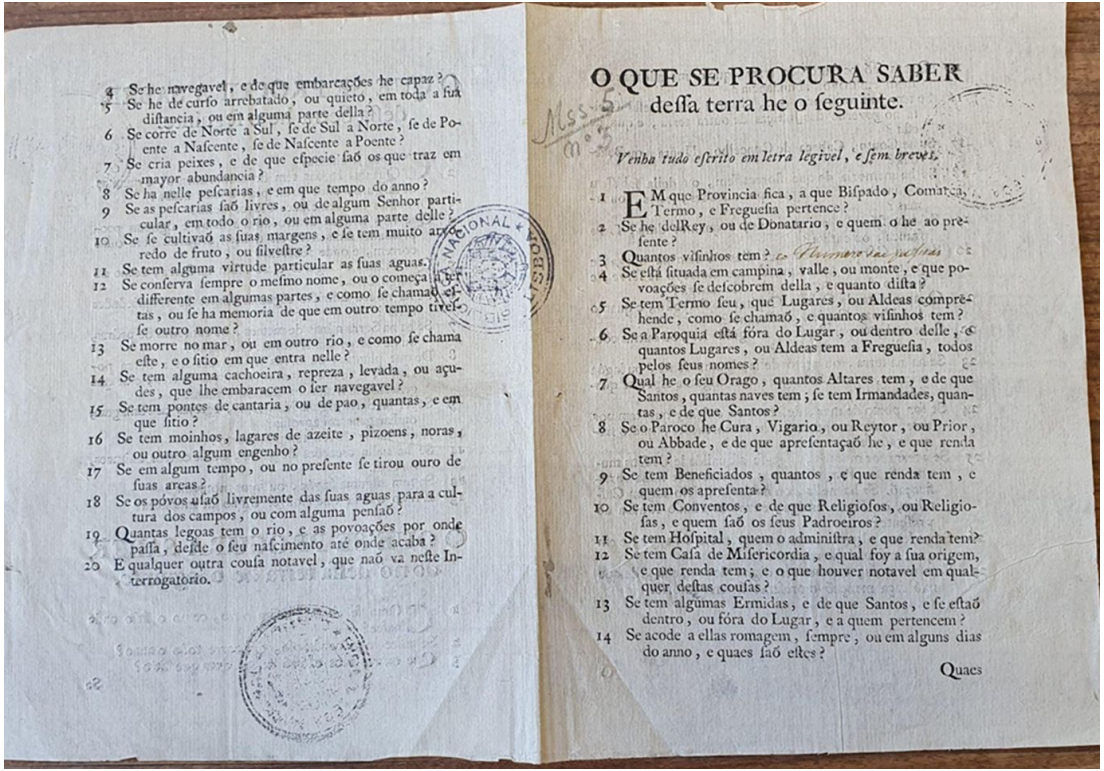
Extração de Eventos
Análise de eventos de conflitos
Anotação de Entidades Nomeadas
Pesquisa por Entidades
Relacionamento Eventos x Entidades

Livro das Monções - Ana Sofia Ribeiro


HISTORIA
DO
FUTURO.
LIVRO
ANTEPRIMEYRO
PROLOGOMENO A TODA A HISTO-
ria do Futuro, em que se declara o fim, & se
provaõ os fundamentos della.
Materia, Verdade, & Utilidades da Historia do Futuro.
ESCRITO PELO PADRE
ANTONIO VIEYRA
da Companhia de JESUS, Prègador de S. Magestade.
LISBOA OCCIDENTAL,
Na Officina de ANTONIO PEDROZO GALRAM.
Com todas as licenças necessarias. Anno de 1718.

Normalização
Análise morfológica e sintática
Elaboração do lexicon
Para breve: anotação semântica

História do Futuro - Ana Paula Banza



4 Se he navegavel, e de que embarcações he capaz?
5 Se he de curſo arrebatado, ou quieto, em toda a ſua diſtancia, ou em alguma parte della?
6 Se corre de Norte a Sul, ſe de Sul a Norte, ſe de Poente a Nafcente, ſe de Nafcente a Poente?
7 Se cria peixes, e de que eſpecie ſaõ os que traz em mayor abundancia?
8 Se ha nelle peſcarias, e em que tempo do anno?
9 Se as peſcarias ſaõ livres, ou de algum Senhor particular, em todo o rio, ou em alguma parte delle?
10 Se ſe cultivaõ as ſuas margens, e ſe tem muito arvoredo de fruto, ou ſilveſtre?
11 Se tem alguma virtude particular as ſuas aguas.
12 Se conſerva ſempre o meſmo nome, ou o começa a ter differente em algumas partes, e como ſe chamaõ eſtas, ou ſe ha memoria de que em outro tempo tiveſſe outro nome?
13 Se morre no mar, ou em outro rio, e como ſe chama eſte, e o ſitio em que entra nelle?
14 Se tem alguma cachoeira, repreza, levada, ou açudes, que lhe embaracem o ſer navegavel?
15 Se tem pontes de cantaria, ou de pao, quantas, e em que ſitio?
16 Se tem moinhos, lagares de azeite, pizoens, noras, ou outro algum engenho?
17 Se em algum tempo, ou no preſente ſe tirou ouro de ſuas areas?
18 Se os póvos uſaõ livremente das ſuas aguas para a cultura dos campos, ou com alguma penſaõ?
19 Quantas legoas tem o rio, e as povoações por onde paſſa, deſde o ſeu naſcimento até onde acaba?
20 E qualquer outra couſa notavel, que naõ va neſte Interrogatorio.

O QUE SE PROCURA SABER
deſſa terra he o ſeguinte.

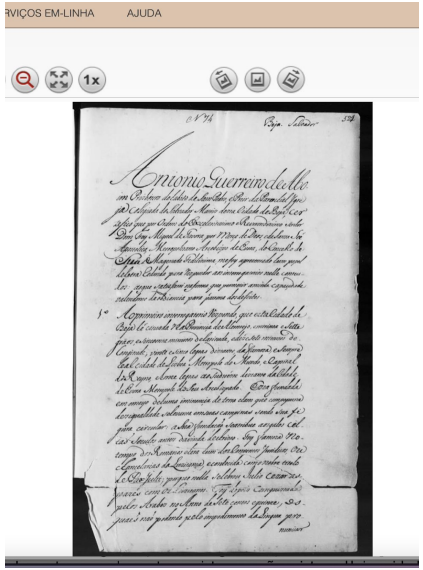
Venha tudo eſcrito em letra legivel, e ſem breves.

1 EM que Provincia fica, a que Biſpado, Comarca, Termo, e Freguefia pertence?
2 Se he delRey, ou de Donatario, e quem o he ao preſente?
3 Quantos viſinhos tem?
4 Se eſtá ſituada em campina, valle, ou monte, e que povoações ſe deſcobrem della, e quanto diſta?
5 Se tem Termo ſeu, que Lugares, ou Aldeas comprehende, como ſe chamaõ, e quantos viſinhos tem?
6 Se a Paroquia eſtá fóra do Lugar, ou dentro delle, e quantos Lugares, ou Aldeas tem a Freguefia, todos pelos ſeus nomes?
7 Qual he o ſeu Orago, quantos Altares tem, e de que Santos, quantas naves tem; ſe tem Irmandades, quantas, e de que Santos?
8 Se o Paroco he Cura, Vigario, ou Reytor, ou Prior, ou Abbade, e de que apreſentaçaõ he, e que renda tem?
9 Se tem Beneficiados, quantos, e que renda tem, e quem os apreſenta?
10 Se tem Conventos, e de que Religioſos, ou Religioſas, e quem ſaõ os ſeus Padroeiros?
11 Se tem Hoſpital, quem o adminiſtra, e que renda tem?
12 Se tem Caſa de Miſericordia, e qual foy a ſua origem, e que renda tem; e o que houver notavel em qualquer deſtas couſas?
13 Se tem algumas Ermidas, e de que Santos, e ſe eſtaõ dentro, ou fóra do Lugar, e a quem pertencem?
14 Se acode a ellas romagem, ſempre, ou em alguns dias do anno, e quaes ſaõ eſtes?

Quaes

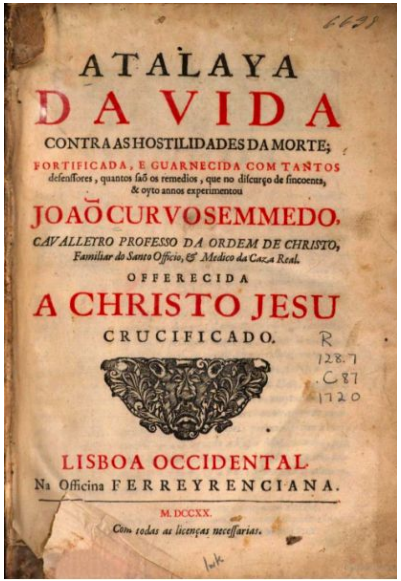
Memórias Paroquiais - Fernanda Olival

BNP, Ms. 5, nº 3



Digitalização (manual)
Normalização (manual)
Anotação de Entidades Nomeadas
Reconhecimento de Entidades Nomeadas (BerTimbau, Albertina)
para expandir anotação para toda a coleção
Análise da fonte a partir de Entidades

Memórias Paroquiais - Fernanda Olival

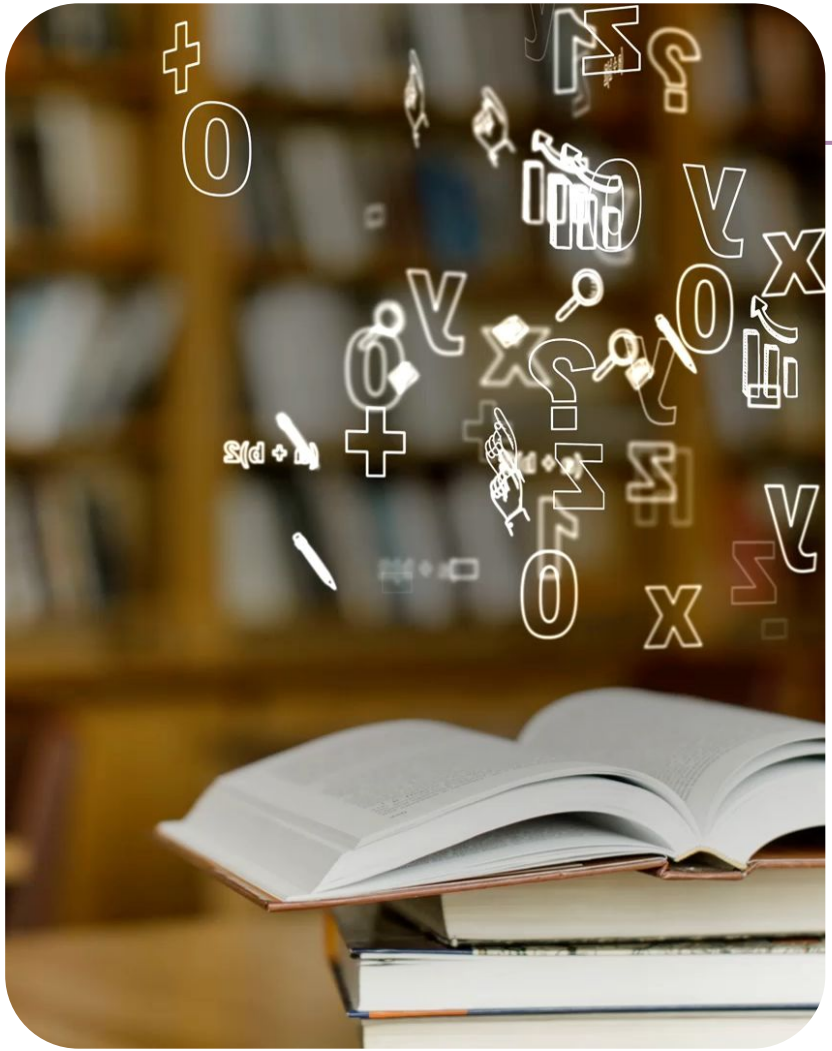


Mapear as antigas terminologias de Saúde em seus equivalentes da atualidade.

Estabelecer quais as áreas da Medicina, da Anatomia e que tipo de doenças e medicamentos eram conhecidos daquela época.

Terminologias usadas nos manuais de Curvo Semedo contrastadas com o vocabulário de Enfermagem

Textos Médicos Séc XVIII - Maria José Finatto



Considerações Finais

PLN e HD

uma correlação de crescimento mútuo

Considerações Finais

É preciso integrar:

- os objetivos de um pesquisador da área de humanidades, como linguística e literatura, ciências sociais, história,
- os objetivos de um pesquisador em PLN de encontrar, aplicar, desenvolver soluções apropriadas para cada tipo de problema, maximizando a correção e a performance das soluções encontradas de acordo com a disponibilidade de recursos.

Celebrando 3 edições do **Workshop de PLN e Humanidades Digitais** junto ao Propor

Considerações Finais

A tecnologia faz parte da definição do **ser humano**

- Quem produz tecnologia deve ter uma formação humanística
- Atenta às origens e consequências do que produz
- Por exemplo a ética, mas não só

Computação não pode ser um mistério acessível a poucos (como a leitura no início da imprensa)

Humanidades, como formação deve incluir conhecimentos de computação

https://brasileiraspln.com/livro-pln/2a-edicao/parte-dominios/cap-humanidades-digitais

Obrigada!

Elementos úteis