
A extração de termos com o Wordsmith Tools a partir de corpora especializados

Stella E. O. Tagnin – USP

UFSCar – 29/11/ 2011

Resumo

- Construção de corpus especializado
 - Extração de termos
 - Identificação de equivalentes
-

Construção do corpus (1)

- Delimitação da área de estudo
 - Monolíngue ou bilíngue?
 - Coleta de textos:
 - corpus comparável: originais em ambas as línguas
 - corpus paralelo: originais e respectivas traduções
 - Meio: internet, impresso?
 - impresso → escanear
 - Tipologia: um tipo ou vários?
 - Fontes: uma ou várias?
 - Autores: um ou vários?
 - Tamanho do texto: trecho ou completo?
-

Construção do corpus (2)

- Limpeza dos textos: imagens, gráficos, bibliografia
- Guardar uma cópia no formato original: HTML.
Word
- Inserção de cabeçalho: metadados
 - toda informação que puder ser útil para outros pesquisadores também:
 - autor
 - fonte
 - data e local de publicação
 - domínio
 - subdomínio
 - tipo de texto

Cabeçalho geral

- <Header>
 - <title>
 - <filename> **Adv05red001** </filename>
 - </title>
 - <author>
 - <name> **Mariazinha da Silva** </name>
 - </author>
 - <sourceText>
 - <language> **inglês** </language>
 - <mode> **written** </mode>
 - <publisher> **Internet** </publisher>
 - <pubPlace> **http://www.xyz.com** </pubPlace>
 - <date> **September 2, 2008** </date>
 - <copyright> [copyrights holder] </copyright>
 - </sourceText>
- </Header>

Cabeçalho de culinária

- <header>
 - <title>
 - <tit> título da receita </tit>
 - <filename> nome do arquivo (siglas) </filename>
 - <collection> colocar "pratos", "doces", "salgadinhos" ou "bebidas" </collection>
 - </title>
 - <sourcetext>
 - <language> PB (português brasileiro); PT (português traduzido); PE (português europeu); IT (inglês traduzido); IO (inglês original) </language>
 - <mode> internet, revista, livro, etc. </mode>
 - <name> nome da fonte, seção </name>
 - <author> nome do autor da receita, se houver </author>
 - <publisher> nome do site ou da empresa que o mantém </publisher>
 - <pubplace> endereço da receita na Internet, lugar onde foi publicado o livro, revista, jornal </pubplace>
 - <pubdate> data de publicação do livro, revista, jornal ou quando o site fornecer a data, que é diferente de data de acesso </pubdate>
 - <accessdate> data da coleta, para arquivos da Internet </accessdate>
 - <comments> comentários do pesquisador </comments>
 - </sourcetext>
 - <researcher> nome do pesquisador que levantou os dados </researcher>
 - </header>
-

Receita etiquetada

<titrec> Peixada com Legumes e Pirão **</titrec>**

<ingr> INGREDIENTES:

1 kg peixe em posta

Cebola, pimentão e tomate

Alho

Cheiro verde

1/2 lata de creme de leite

Água

</ingr>

<modFaz> Preparo:

Temperar as postas com limão, sal, pimenta e alho. Refogar os temperos com óleo e um pouco de azeite de oliva. Juntar o peixe, o molho em que foi temperado e o repolho. Acrescentar água fervendo, um pouco de sal e colorau e deixar cozinhar. Quando o peixe estiver pronto, juntar o creme de leite, as batatas e os ovos cozidos. Com uma parte do caldo fazer um pirão.

</modFaz>

<class>Fácil **</class>**

<coment>Se quiser uma versão mais light, elimine o creme de leite. **</coment>**

Construção do corpus (3)

- Nomeação dos arquivos: mnemônico, domínio, língua
 - Informática: InfoIngAc01
 - Culinária: ppa01IOA; aco03POB
 - Gravação em .txt
 - Estruturação dos arquivos
-

- Corpora
 - Austrália Completo
 - Australia ptg
 - Bone quality
 - Chico
 - CorRefIng
 - CorRefPtg
 - CorTec
 - Culinaria
 - Culinária Gourmand
 - Foreign Policy
 - Fotografia
 - Hipertensão
 - Informática
 - Investments
 - Journalism
 - Financial Times
 - Los Angeles Times
 - Business
 - Entertainment
 - Front Page
 - Health
 - Nation and World
 - SPORTS
 - TRAVEL
 - The Guardian
 - The New York Times
 - The Times
 - USA Today
 - Washington Post
 - Journalism Politics
 - Kidney corpus
 - LW Parcial Geral
 - Lw Ref Total
 - MAC-MORPHO_PESQUISA
 - Mercosul
 - Mormon
 - NatGeo
 - nefrologia
 - Núcleo José Reis
 - NVT F

- 171000A
Documento de texto
18 KB
- 031099B
Documento de texto
13 KB
- 031099E
Documento de texto
7 KB
- 031099H
Documento de texto
14 KB
- 081000A
Documento de texto
30 KB
- 121100A
Documento de texto
29 KB
- 151000A
Documento de texto
20 KB
- 221000A
Documento de texto
19 KB
- 270800A
Documento de texto
19 KB
- 301200A
Documento de texto
27 KB

- 031000A
Documento de texto
19 KB
- 031099C
Documento de texto
27 KB
- 031099F
Documento de texto
8 KB
- 051100A
Documento de texto
19 KB
- 081000B
Documento de texto
19 KB
- 121100B
Documento de texto
19 KB
- 171200A
Documento de texto
19 KB
- 240900A
Documento de texto
20 KB
- 291000A
Documento de texto
19 KB
- 301200B
Documento de texto
18 KB

- 031099A
Documento de texto
7 KB
- 031099D
Documento de texto
7 KB
- 031099G
Documento de texto
6 KB
- 060800A
Documento de texto
18 KB
- 101000A
Documento de texto
18 KB
- 130800A
Documento de texto
19 KB
- 200800A
Documento de texto
18 KB
- 241200A
Documento de texto
19 KB
- 300700A
Documento de texto
18 KB
- 301200C
Documento de texto
13 KB

Textos

Construção automática

- BootCat: <http://bootcat.sslmit.unibo.it>
 - BootCat tutorial (5 partes):
http://bootcat.sslmit.unibo.it/wiki/doku.php?id=tutorials:basic_1
 - Bing Appld: <http://www.bing.com/developers>
 - Seed words → tuples → URLs → seleção → finalização
 - Vantagem: rapidez, volume de textos
 - Desvantagem: pouco controle, um grande arquivo sem subdivisões
-

Extração de termos

- Lista de palavras – WordList
 - Palavras-chave – Keywords
 - Seleção de candidatos a termo
 - Concordâncias
-

N	Word	Freq.	%	Texts	%_lemmas	Set
1	DE	10.283	5,55	50	100,00	
2	#	9.863	5,32	50	100,00	
3	A	6.343	3,42	50	100,00	
4	E	4.586	2,47	50	100,00	
5	O	3.335	1,80	50	100,00	
6	DA	3.045	1,64	50	100,00	
7	EM	2.853	1,54	50	100,00	
8	DO	2.676	1,44	50	100,00	
9	QUE	2.478	1,34	50	100,00	
10	COM	2.247	1,21	50	100,00	
11	É	1.587	0,86	49	98,00	
12	PACIENTES	1.410	0,76	49	98,00	
13	PARA	1.384	0,75	50	100,00	
14	NA	1.308	0,71	50	100,00	
15	OS	1.299	0,70	50	100,00	
16	RENAL	1.273	0,69	50	100,00	
17	POR	1.189	0,64	49	98,00	
18	OU	1.098	0,59	49	98,00	
19	NO	1.072	0,58	50	100,00	
20	UM	1.025	0,55	50	100,00	
21	DOS	1.019	0,55	48	96,00	
22	NÃO	906	0,49	48	96,00	
23	UMA	893	0,48	50	100,00	
24	COMO	775	0,42	49	98,00	
25	AS	754	0,41	49	98,00	
26	SER	732	0,39	48	96,00	
27	SE	729	0,39	47	94,00	
28	À	709	0,38	49	98,00	
29	MAIS	647	0,35	48	96,00	
30	SÃO	634	0,34	48	96,00	
31	FOI	506	0,27	47	94,00	

Lista de palavras

Key word	Freq.	%RC. Freq	RC. %	Keyness	P	lemmas	Set
PACIENTES	1.410	0,76	85	3.750,48	0000000		
RENAL	1.273	0,69	3	3.629,38	0000000		
DIÁLISE	469	0,25	0	2.455,93	0000000		
INSUFICIÊNCIA	431	0,23	15	2.127,87	0000000		
CRÔNICA	369	0,20	15	1.807,68	0000000		
HEMODIÁLISE	337	0,18	2	1.740,27	0000000		
ARTERIAL	299	0,16	0	1.565,47	0000000		
PACIENTE	304	0,16	84	1.198,96	0000000		
HIPERTENSÃO	218	0,12	3	1.109,99	0000000		
PERITONEAL	192	0,10	0	1.005,15	0000000		
TRANSPLANTE	202	0,11	6	1.004,04	0000000		
IRC	183	0,10	0	958,02	0000000		
#	9.863	5,32	91.002	3,87	870,59	0000000	
RENAIS	164	0,09	0	858,54	0000000		
TRATAMENTO	401	0,22	604	0,03	839,18	0000000	
GLOMERULAR	152	0,08	0	795,71	0000000		
PATIENTS	146	0,08	0	764,30	0000000		
RINS	149	0,08	4	743,56	0000000		
INGESTÃO	151	0,08	10	717,05	0000000		
CSA	138	0,07	1	710,70	0000000		
DOSE	149	0,08	15	681,94	0000000		
SEXO	162	0,09	32	679,18	0000000		
DOENÇA	237	0,13	197	672,66	0000000		
CAPD	128	0,07	0	670,06	0000000		
URÉIA	143	0,08	12	665,96	0000000		
DL	130	0,07	2	660,10	0000000		
SÓDIO	159	0,09	35	654,52	0000000		
PRESSÃO	260	0,14	294	0,01	639,83	0000000	
CREATININA	122	0,07	0	638,64	0000000		
HD	130	0,07	8	620,65	0000000		
ANEMIA	118	0,06	1	606,30	0000000		

Wordlist vs LW agrarias+exatas +ciências sociais

Key word	Freq.	%RC.	Fred	RC. %	Keyness	P	lemmas	Set
RENAL	1.273	0,69	25		4.443,60	0000000		
PACIENTES	1.410	0,76	556	0,06	3.036,73	0000000		
DIÁLISE	469	0,25	4		1.678,48	0000000		
INSUFICIÊNCIA	431	0,23	39		1.328,22	0000000		
#	9.863	5,32	34.395	3,52	1.261,91	0000000		
HEMODIÁLISE	337	0,18	6		1.179,65	0000000		
CRÔNICA	369	0,20	33		1.138,88	0000000		
ARTERIAL	299	0,16	47		839,71	0000000		
PERITONEAL	192	0,10	0		705,21	0000000		
IRC	183	0,10	0		672,15	0000000		
HIPERTENSÃO	218	0,12	41		588,68	0000000		
GLOMERULAR	152	0,08	0		558,27	0000000		
RENAIS	164	0,09	9		534,74	0000000		
CSA	138	0,07	0		506,84	0000000		
URÉIA	143	0,08	3		497,00	0000000		
DL	130	0,07	0		477,45	0000000		
CAPD	128	0,07	0		470,10	0000000		
CÁLCIO	163	0,09	24		463,68	0000000		
PRESSÃO	260	0,14	155	0,02	460,38	0000000		
HD	130	0,07	2		457,42	0000000		
RINS	149	0,08	14		456,61	0000000		
CREATININA	122	0,07	1		436,80	0000000		
TRANSPLANTE	202	0,11	84		424,79	0000000		
NÍVEIS	276	0,15	210	0,02	421,96	0000000		
UREMIA	114	0,06	0		418,68	0000000		
INGESTÃO	151	0,08	26		415,90	0000000		
TRATAMENTO	401	0,22	492	0,05	415,19	0000000		
PACIENTE	304	0,16	278	0,03	407,52	0000000		
MEDLINE	110	0,06	0		403,99	0000000		
DIALÍTICO	109	0,06	0		400,31	0000000		
SÓDIO	159	0,09	40		398,15	0000000		

WordList vs LW saúde + biológicas

plot links clusters filenames notes source text

Type-in SÉRICA

Corpus de Referência

- Qual a diferença?
 - Corpus de língua geral:
 - agrárias + exatas + ciências sociais
 - → termos da medicina em geral
 - Corpus **especializado** mais geral:
 - saúde + biológicas → termos mais específicos
-

KW
Vs
LW saúde

1 RENAL
2 PACIENTES
3 DIÁLISE
4 INSUFICIÊNCIA
5 #
6 HEMODIÁLISE
7 CRÔNICA
8 ARTERIAL
9 PERITONEAL
10 IRC
11 HIPERTENSÃO
12 GLOMERULAR
13 RENAIIS
14 CSA
15 URÉIA
16 DL
17 CAPD
18 CÁLCIO
19 PRESSÃO
20 HD
21 RINS
22 CREATININA
23 TRANSPLANTE
24 NÍVEIS
25 UREMIA
26 INGESTÃO
27 TRATAMENTO
28 PACIENTE
29 MEDLINE
30 DIALÍTICO
31 SÓDIO
32 DOSE
33 NEFROLOGIA
34 SEVELAMER
35 SOBREVIDA
36 RIM
37 ALBUMINA
38 **KG**
39 ANEMIA
40 EXCREÇÃO
41 **PLASMÁTICA**
42 **CISTOS**
43 **SÉRICA**
44 **FUNÇÃO**
45 **SOLUTOS**
46 **URÊMICOS**
47 ENXERTO
48 PATIENTS
49 KIDNEY
50 **CONCENTRAÇÃO**

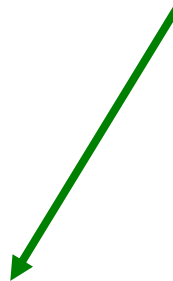
1 PACIENTES
2 RENAL
3 DIÁLISE
4 INSUFICIÊNCIA
5 CRÔNICA
6 HEMODIÁLISE
7 ARTERIAL
8 PACIENTE
9 HIPERTENSÃO
10 PERITONEAL
11 TRANSPLANTE
12 IRC
13 #
14 RENAIIS
15 TRATAMENTO
16 GLOMERULAR
17 PATIENTS
18 RINS
19 INGESTÃO
20 CSA
21 DOSE
22 **SEXO**
23 DOENÇA
24 CAPD
25 URÉIA
26 DL
27 SÓDIO
28 PRESSÃO
29 CREATININA
30 HD
31 ANEMIA
32 UREMIA
33 SOBREVIDA
34 MEDLINE
35 RIM
36 DIALÍTICO
37 **SANGUE**
38 **LINFÓCITOS**
39 NEFROLOGIA
40 CÁLCIO
41 ALBUMINA
42 **VASCULAR**
43 **ANTICORPOS**
44 SEVELAMER
45 **EXCREÇÃO**
46 **MEMBRANA**
47 SOLUTOS
48 ENXERTO
49 NÍVEIS
50 KIDNEY

KW
Vs
LW agrárias

Concordance		Set	Tag	Word #	Sen	Sen
súbitas dos níveis pressóricos dos pacientes. O liquor de urêmicos "normais" mostra pressões elevadas, proteínas em				7.479	31828%	
, catecolaminas e angiotensina II. 6. Ausência de desequilíbrio urêmico. Aliás, observa-se um bem-estar considerável como				9.681	40630%	
no sistema nervoso, culminando com o estado de desequilíbrio urêmico observado após certas hemodiálises muito eficientes ou				1.368	4278%	
tempo, o excesso de PTH contribui para a piora da anemia do urêmico, independentemente dos níveis de eritropoetina (EPO),				1.412	4443%	
pela deterioração mental, ao lado do estado pré-convulsivo do urêmico. Na uremia, devido ao hiperparatireoidismo, o				1.341	4130%	
cérebro e no liquor concorre pela disfunção cerebral e mental do urêmico. Um agente adsorvente, o AST-120, retira-o do intestino e				3.730	13830%	
sido implicada também na patogênese da hipertensão arterial do urêmico. TRATAMENTO Tratamento conservador O curso natural				8.161	35230%	
pela dimetilação da metionina e implicado no risco coronariano do urêmico. O excesso de homocisteína ocorre pelo déficit de ácido				3.871	14430%	
avaliação radiográfica pode demonstrá-la em cerca de 40% dos urêmicos e a análise histológica (biópsia óssea) pode chegar a				6.903	29152%	
e à osteodistrofia, o que incapacita enorme porcentagem dos urêmicos, mesmo quando adequadamente dialisados. Assim, o				1.306	3931%	
urina. Na uremia é retido, deteriorando a função renal residual dos urêmicos via esclerose glomerular. Este soluto no cérebro e no				3.712	13730%	
As prostaglandinas PGE2 e PGI2 estão aumentadas na urina dos urêmicos, provavelmente como fatores que contrabalançam os				2.282	7755%	
na prevenção do balanço nitrogenado negativo prolongado em urêmicos17. NECESSIDADES CALÓRICAS EM CRIANÇAS EM				2.021	7530%	
, sua meia-vida é curta (60 minutos), tanto em normais como em urêmicos. Anticoagulação em hemodiálise Cascata normal de				7.262	30930%	
ascórbico, ácido fólico e de piridoxina. Os níveis de vitamina A em urêmicos geralmente é elevado e sua administração rotineira deve				10.054	42144%	
de vitaminas hidros-solúveis pode ser necessária em urêmicos, principalmente quando em programa de diálise;				10.025	42039%	
o conhecimento da farmacocinética de diversas substâncias em urêmicos e facilitado a tarefa de prescrevê-las com segurança				9.696	41071%	
uso regular de heparina e discrasias eventualmente presentes em urêmicos. Hocken & Marwah, em 1971,4 sugeriram que a perda				284	630%	
1990;335:1409-10. 11. Francheschini N. Citologia hepática em urêmicos portadores crônicos do vírus da hepatite B [tese de				4.165	24433%	
. Outras doenças não-específicas são também vistas em urêmicos crônicos (hemossiderose, esteatose e peliose hepática).				6.768	28450%	
do início do tratamento hemodialítico10, sugerindo que o estado urêmico seja suficiente para induzir a ativação celular. Tendo por				1.943	5837%	
tubulares, decorrente de fatores proliferativos gerados pelo estado urêmico, com a conseqüente obstrução e dilatação da luz tubular				708	1844%	
aderirem a elas. Finalmente, deve-se lembrar que o estado urêmico está acompanhado de anorexia, náuseas e vômitos, o				8.894	38020%	
à deficiência de calcitriol e em parte relacionado ao próprio estado urêmico. Como resultado de todos os fatores assinalados até o				5.253	21130%	
variedade de distúrbios imunológicos que acompanham o estado urêmico. Estas alterações, que serão analisadas neste capítulo,				336	530%	
(IRT) permitiu que se estabelecesse um elo entre o estado urêmico e a presença de imunodepressão. Assim a denominada				139	037%	
realmente alteradas pela presença de fatores séricos no estado urêmico. Finalmente, diversos estudos têm evidenciado também				3.058	9030%	
A função fagocitária de polimorfonucleares (PMN) no estado urêmico tem sido objeto de vários estudos nos últimos anos23'4.				2.299	6955%	
do oxigênio e liberação de enzimas microbicidas26. No estado urêmico, as funções dos PMN avaliadas in vitro em geral estão				2.665	8127%	
, aqueles com acidentes vasculares cerebrais, os extremamente urêmicos, os com hipertensão grave ou maligna etc. Para estes				8.116	34074%	
para o crescimento de insuficiência cardíaca congestiva. Estes urêmicos, como os depressores micóidicos dialisados,				6.388	26419%	

Cluster	Freq.	Set	Length	Related
EM PACIENTES URÊMICOS	19		3	} paciente(s) urêmico(s)
DOS PACIENTES URÊMICOS	10		3	
DO PACIENTE URÊMICO	6		3	
URÊMICOS DIALISADOS OU	5		3	} paciente(s) urêmico(s) dialisado(s)
PACIENTES URÊMICOS DIALISADOS	5		3	
O PACIENTE URÊMICO	5		3	

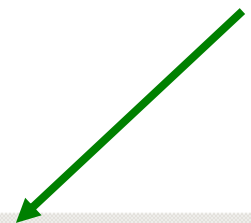
Clusters - agrupamentos



Word	With	Relation	Texts	Total	Total Left	Total Right	L5	L4	L3	L2	L1	Centre	R1	R2	R3
URÊMICOS	urêmicos	0,000	12	86	0	0	0	0	0	0	0	86	0	0	0
URÊMICO	urêmico	0,000	9	56	0	0	0	0	0	0	0	56	0	0	0
PACIENTES	urêmicos	0,000	10	43	43	0	0	0	0	1	42	0	0	0	0
EM	urêmicos	0,000	10	41	34	7	2	2	0	23	7	0	2	2	1
DE	urêmicos	0,000	8	31	17	14	2	5	2	7	1	0	1	3	2
DOS	urêmicos	0,000	7	20	19	1	0	1	0	14	4	0	0	0	1
DO	urêmico	0,000	6	18	16	2	1	1	0	9	5	0	0	1	0
O	urêmico	0,000	5	18	16	2	0	2	1	12	1	0	0	1	1
NO	urêmico	0,000	5	16	14	2	0	1	0	11	2	0	0	0	0
PACIENTE	urêmico	0,000	6	14	14	0	0	0	0	0	14	0	0	0	0
OS	urêmicos	0,000	4	13	8	5	0	0	0	8	0	0	1	1	1
NOS	urêmicos	0,000	6	11	10	1	0	0	1	3	6	0	0	0	0
ESTADO	urêmico	0,000	4	11	11	0	0	0	2	0	9	0	0	0	0
SOLUTOS	urêmicos	0,000	2	9	8	1	0	0	0	0	8	0	0	0	1
SINTOMAS	urêmicos	0,000	4	8	7	1	0	0	0	0	7	0	0	0	1
DIALISADOS	urêmicos	0,000	3	7	0	7	0	0	0	0	0	0	5	1	0
RATOS	urêmicos	0,000	2	5	5	0	0	0	0	0	5	0	0	0	0
PLASMA	urêmico	0,000	2	5	5	0	0	0	0	1	4	0	0	0	0

Busca: urêmico*

colocados



A identificação de equivalentes tradutórios em corpora comparáveis

Equivalência em tradução

Definição pragmática:

- um termo que “funcione” no texto de chegada como “funciona” no texto de partida.

O tradutor quer

- tradução fluente a fim de garantir melhor entendimento por parte do leitor.

Para isso não deveria

- empregar um termo não usual que, com certeza, causaria certo estranhamento nesse leitor.
-

Dicionários

- falta de critério na compilação dos termos
 - falta de exemplos de uso
 - falta de atualização devido ao rápido avanço das pesquisas científicas e tecnológicas.
 - falta de cobertura de diversas áreas técnicas
-

Vantagens de um corpus

- pode ser construído de acordo com as necessidades do tradutor/ pesquisador/ terminólogo/ produtor de texto
 - pode ser constantemente atualizado
 - fornece exemplos autênticos de uso
 - portanto, confere segurança ao tradutor na escolha do termo a empregar.
-

Instrumentos Contratuais

- lista de palavras em ambas as línguas
- português: **contrato** era a palavra de conteúdo de maior frequência, com **1832** ocorrências
- *contract*: apenas **186** vezes

contrato ≠ contract

- em inglês: a palavra de conteúdo mais frequente era **agreement**, com **1724** ocorrências.
-

agreement vs. contrato

1a For the purposes of this Agreement, all merchantable Logs 6" in diameter

1b Constitui objeto do presente contrato o intercâmbio eletrônico de documentos

2a 12.2 This Agreement may be cancelled by either party, at it

2b 13.2 A rescisão deste contrato implicará retenção de créditos decorren

3a The term of this Agreement shall expire June 30, 2001, (the "Term"

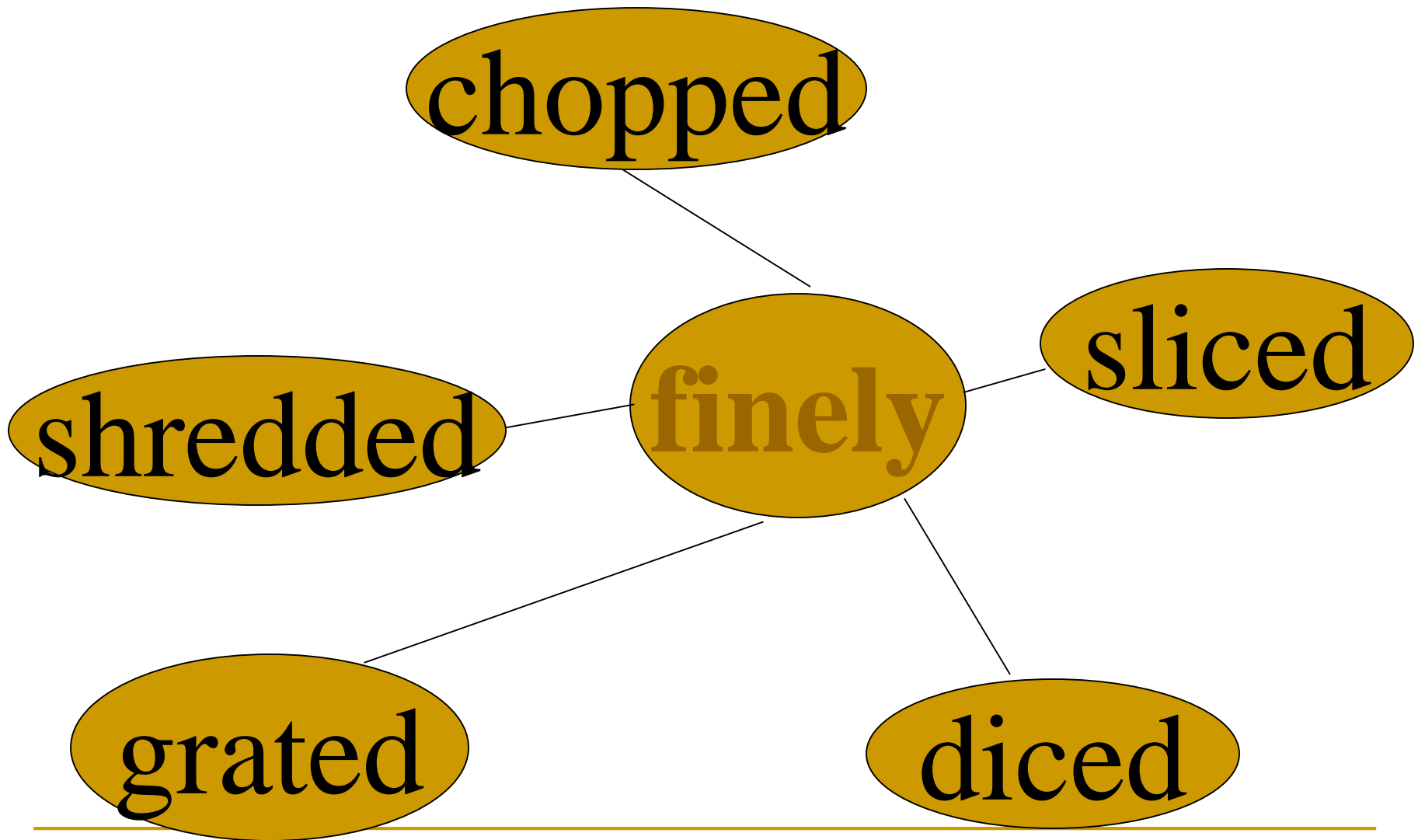
3b O presente contrato terá prazo de (xxx), iniciando-se no di

Adhesion contract (Google)

42.000 "adhesion contract"

890 "adhesion agreement" (maioria de sites traduzidos)

Culinária



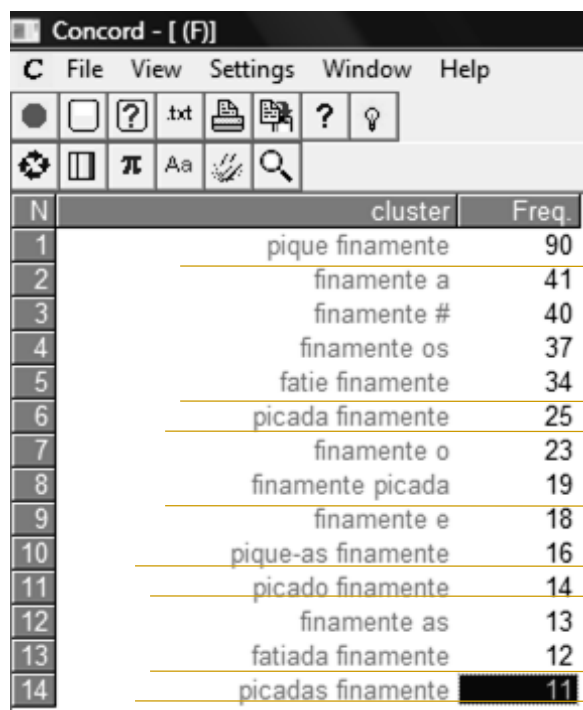
Obviamente

Finely = finamente



Busca por equivalentes

- finely (2.987 oc.) = finamente (337 oc.)?



The screenshot shows a window titled "Concord - [(F)]" with a menu bar (File, View, Settings, Window, Help) and a toolbar. Below the toolbar is a table with three columns: "N", "cluster", and "Freq.". The table lists 14 clusters of the word "finamente" and their frequencies.

N	cluster	Freq.
1	pique finamente	90
2	finamente a	41
3	finamente #	40
4	finamente os	37
5	fatie finamente	34
6	picada finamente	25
7	finamente o	23
8	finamente picada	19
9	finamente e	18
10	pique-as finamente	16
11	picado finamente	14
12	finamente as	13
13	fatiada finamente	12
14	picadas finamente	11



clusters de “finamente”

- **Não-equivalência**
(finely ≠ finamente)

- **Método indireto:**
chop = picar

- **colocados** de “pique/picad*”



Finely chopped = ???

- Soy sauce - 1 tbsp Onion - 1 medium, **finely chopped** Celery - 3 sticks, finely chopp
- apples - 450g (1 lb), peeled, cored and **finely chopped** Onions - 225g (8 oz), finely ch
- Milk - 600 ml (1 pint) Onion - 2 tbsp, **finely chopped** Celery - 2 tbsp, finely chopped

chop = picar

- **29** ocorrências em que o colocado é “fino” ou algum derivado seu:

“picad* fin*” (**18** ocorrências),
“finamente picad*” (**11** ocorrências).

Resultados para picad*

- o advérbio com que “picad*” mais co-ocorre é “bem” (79 ocorrências): “bem picada, bem picado” etc.
 - “picadinh*” (96 ocorrências, sendo 10 de “bem picadinha”).
 - melhores equivalências para *finely chopped* seriam **“bem picad*”** ou **“picadinh*”**
-

Resultados para picad*

- 2 cebolas médias **bem picadas**
 - ½ dente de alho **bem picado**
 - junte os tomates pelados **bem picados.**
 - Calabresa **picadinha**
 - 100 g de bacon **picadinho**
 - 2 dentes de alho **picadinhos**
 - Polvilhar salsa **bem picadinha**
 - ½ cebola **bem picadinha**
-

finely sliced

Slice = cortar em fatias (?* fatiar)

- Calda 4 laranjas descascadas **cortadas em fatias finas**
- 200 g de cebola **cortada em fatias finas**
- 1 pepino sem sementes **cortado em fatias finas**
- 6 rabanetes, **cortados em fatias finas**
- Juntar as batatas **cortadas em fatias finas**.
- Decore a quiche com um alho-poró cru **cortado em rodela finas**.
- 1 cebola média **cortada em rodela finas**
- 400 g de lingüiça portuguesa **cortada em rodela finas**

finely diced

- 50 g de bacon **em cubinhos**
 - 500 g de peito de frango cozido e **cortado em cubinhos**
 - 1 tomate sem pele e sementes **cortado em cubinhos**
 - 1 abacaxi médio **cortado em cubinhos**
 - 150 g de presunto cozido **cortado em cubinhos**
 - 1/2 xícara (chá) de queijo prato **cortado em cubinhos**
 - 1 cebola grande, **cortada em cubinhos**
 - 100 g de bacon **em cubos pequenos**
 - 300g de abóbora moranga **cortada em cubos pequenos**
 - 200 g de bacon **cortado em cubos pequenos**
-

finely grated

2 ocorrências para “queijo”

- 2 col. (sopa) de queijo parmesão **ralado fino**
- 80g de queijo gruyère **ralado fino**

Default

- 2 colheres (sopa) de parmesão **ralado**
- 50 g de queijo parmesão, **ralado**

32 ocorrências para “queijo” + “grosso”

- 1 xícara de queijo prato **ralado grosso**
- 2 xícaras de queijo mussarela **ralado grosso** (200 g)
- Para polvilhar 50 g de queijo parmesão **ralado grosso**

Hortaliças e chocolate

- 1 cebola **ralada fino**
 - 4 xícaras de repolho **ralado fino**
 - Cobertura de chocolate **ralado bem fino**
-

Hipertensão Arterial

heart/coração – cardi-

- *Heart* : 768 ocorrências
- heart failure & heart disease
- Tradução de *heart failure*??

10
11
hows

dy fou 154 th ISH reduced the incidence of stroke, **heart failure**, and
si 255 AG increases with advanced age, stroke, **heart failure**, s
had 706 ol subjects, 12 and patients with severe **heart failure**.
ysfunctio 281 I infarction or in patients with severe **heart failure**. These
inuria 365); acute pulmonary edema, congestive **heart failure**, left
e was si 312 ocardial infarction, stroke, congestive **heart failure**,
495 h hypertension, obesity, and congestive **heart failure**, the

Hipertensão Arterial

heart/coração – cardi-

- *Heart* : 768 ocorrências
- heart failure & heart disease

7 dy fou 154 th ISH reduced the incidence of stroke, **heart failure**, and 8
 si 255 AG increases with advanced age, stroke, **heart failure**, 9 as
had 706 ol subjects, 12 and patients with severe **heart failure**. 18
10 ysfunctio 281 I infarction or in patients with severe **heart failure**. These
inuria 365); acute pulmonary edema, **congestive heart failure**, left 12
11 e was si 312 ocardial infarction, stroke, **congestive heart failure**,
hows 495 h hypertension, obesity, and **congestive heart failure**, the

- mHpresença de **insuficiência cardíaca congestiva**, hemorragia cere-2
- is quando existe **insuficiência cardíaca congestiva** associada. Podem causar
- s o severa ou **insuficiência cardíaca congestiva** associada. Limitações do

Copa do Mundo

■ marcar um gol

■ 39 . Don't blame us that they didn't **score a goal.**" Instead, blame the defending champions, they didn't **score a goal** and went home whether the Americans can finally **score a goal** or two on their

■ gol contra

■ 1 defender Fabio Grosso, who scored the go-ahead **goal against** Germany

■ 2 le in this World Cup for Trinidad. England's first **goal against** T&T came

■ 3 tional speed. But with the exception of the lovely **goal against** Ghana that

■ Os croatas começaram na frente, com um **gol contra** de Pablo Ibañez, 5
Azzurra só foi vazada uma única vez: um **gol contra** do lateral Zaccardo,

gol contra = own goal

- 4 Italy defender **Cristian Zaccardo** scored an **own-goal** 27 minutes i
 - 5 by a reserve defender, **Cristian Zaccardo**, who put the ball into his own net
 - 6 nced out by an **own goal** from **Cristian Zaccardo**, but the real talking point of a
 - 7 an **own goal** from **Cristian Zaccardo** put the Americans back on level
-
- 4 imitation of his team in the 1998 Cup by an **own goal**, was assassinated when he
 - 5 minutes for U.S. soccer there -- we had an **own goal** and a red card,' " said
 - 7 nder, Andres Escobar, scored an **own goal** to help the United States to a
 - 8 araguay captain Carlos Gamarra scores an **own goal** to give England the lead as
 - 9 en up one goal in six games, and it was an **own goal**. And the Italians have done
 - 10 wanted." England had to rely on an early **own goal** by Paraguay captain Carlos
 - 11 the interval 1-0 in front thanks to an early **own goal** by Carlos Gamarra. It took
 - n repaired, after the three red cards and the **own goal** and the disallowed goal,
 - 15 ent was overpowered from the onset. The **own goal** only made it worse, then in
 - 16 es three weeks ago on Cristian Zaccardo's **own goal**. The Azzurri had been
-

Como identificar equivalentes?

1. Lista de palavras → palavras de maior frequência: **contrato – agreement;**
 2. Colocados da palavra de busca: marcar um **gol** → score a **goal**
 3. Tradução dos colocados: **finely** + chopped/diced/sliced/grated/shredded → picad*, cortad*, ralad*
 4. Colocado cognato: congestive heart failure → insuficiência cardíaca **congestiva**
 5. Contexto: gol contra → **Zaccardo, Aloisi**
-

Obrigada

Stella

seotagni@usp.br
