



# CETEMPúblico

Um grande corpus de  
linguagem jornalística portuguesa

Paulo Rocha & Diana Santos

Processamento computacional do português

SINTEF Telecomunicações e Informática

**C**  
**E**  
**T**  
**E**  
**M**

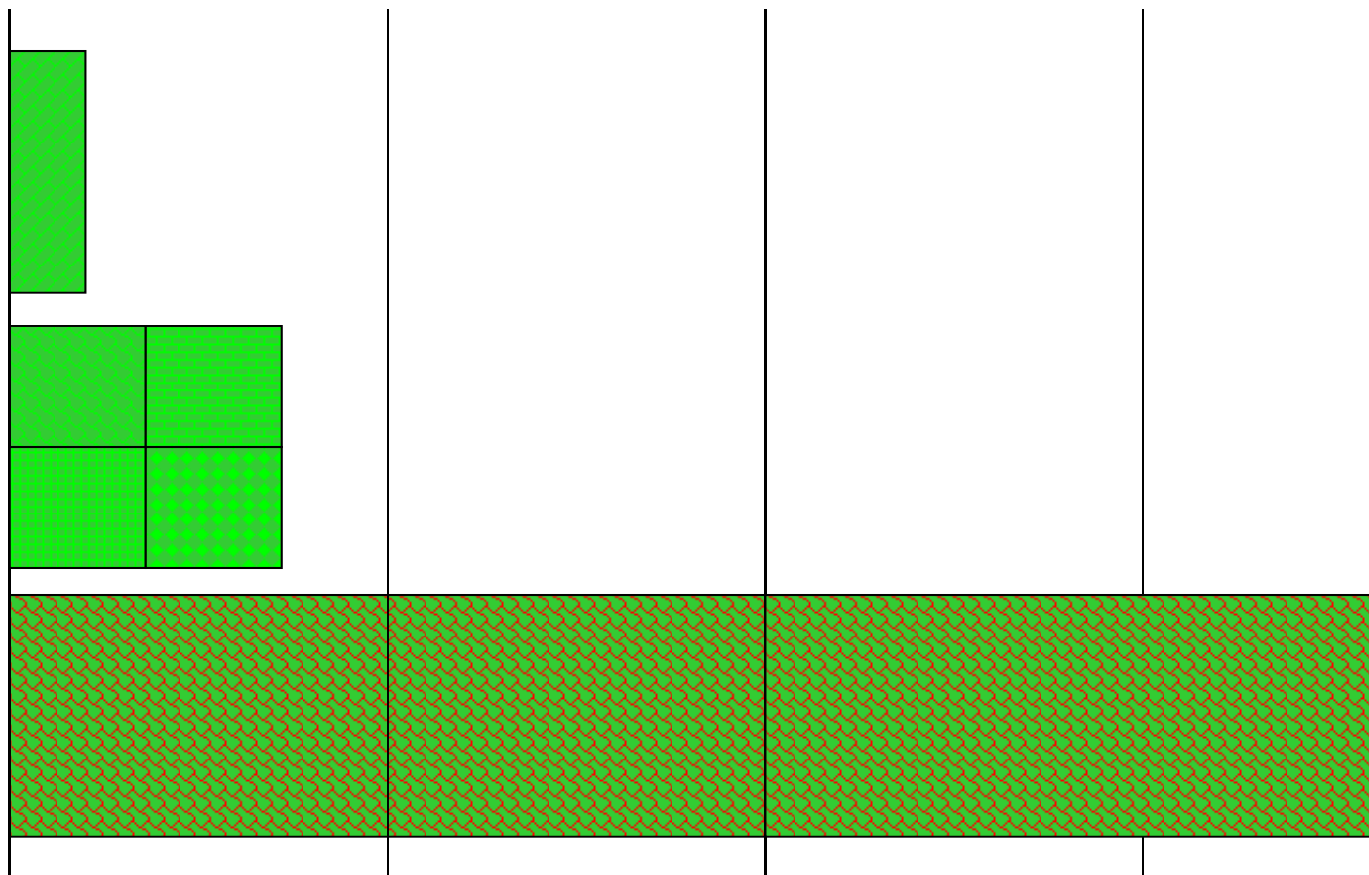
**Público**



Outros corpora

NILC/São Carlos

CETEMPúblico

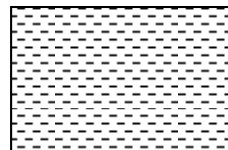


**C**orpus de  
**E**xtractos de  
**T**extos  
**E**lectrónicos  
**M**CT /  
**P**úblico



X

1500



X

1.500.000

Palavra

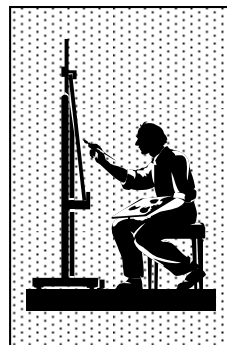
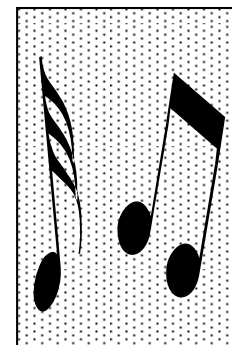
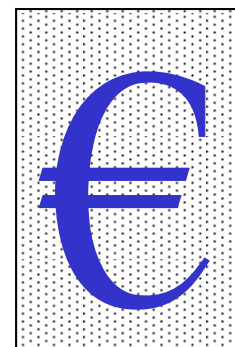
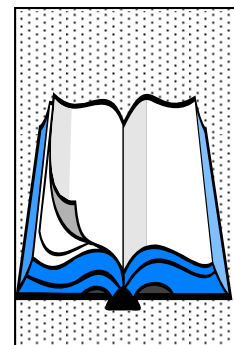
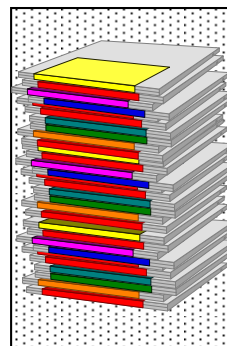
X

180.000.000



# Público

1			



C  
E  
T  
E  
M

# Público

**Público**  
**Lisboa**



**Público**  
**Porto**



# Público

Guardian THE IRISH TIMES Libération

La Presse TO BHMÁ EL PAÍS

FOLHA DE S. PAULO

**MLADINA**

LE SOIR

der Standard

Aftenposten

יָרָאָה

Независимая Газета

LA STAMPA Süddeutsche Zeitung

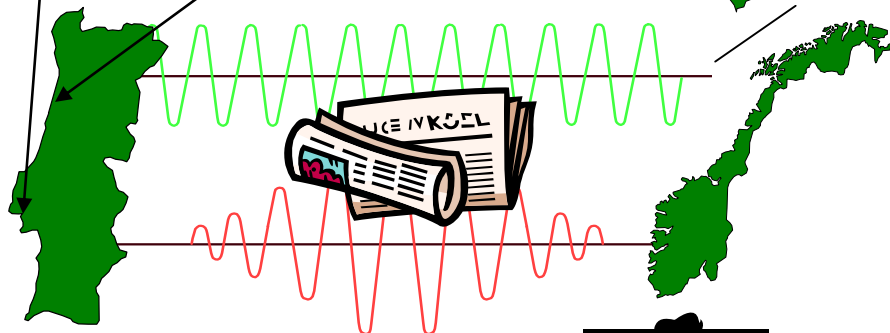
LIDOVÉ NOVINY

BerlingskeTidende



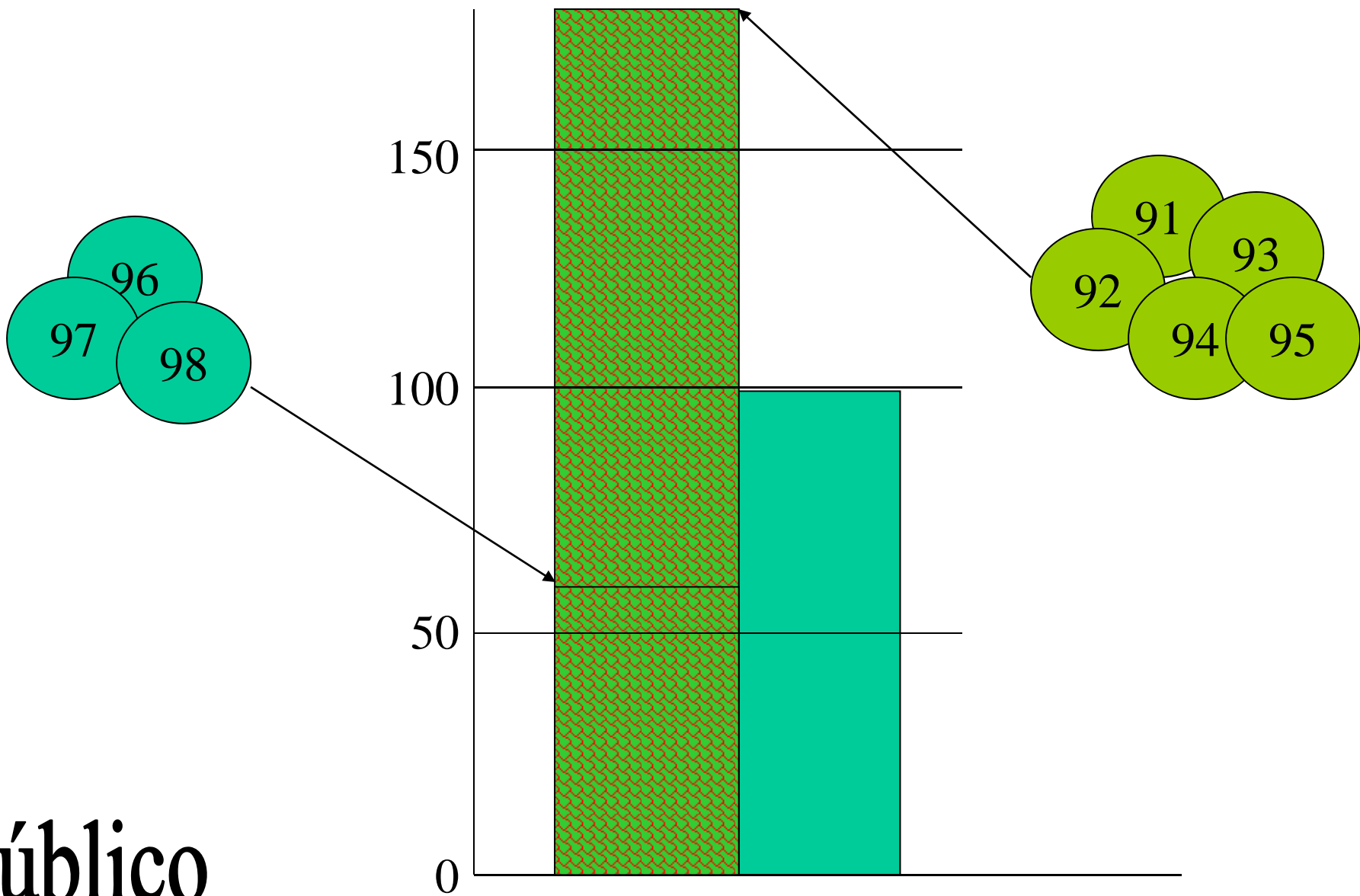
**C**  
**E**  
**T**  
**E**  
**M**

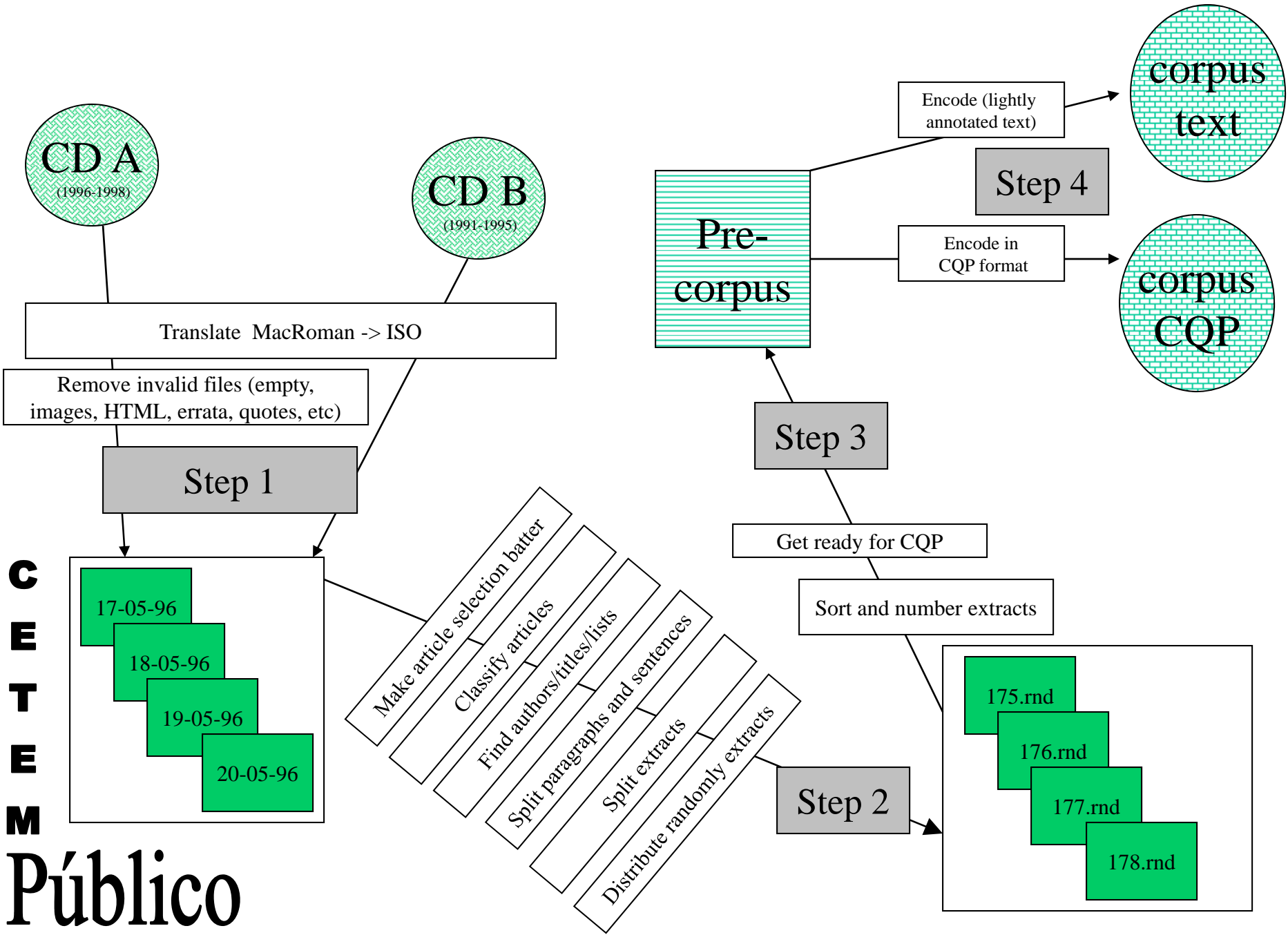
**Público**





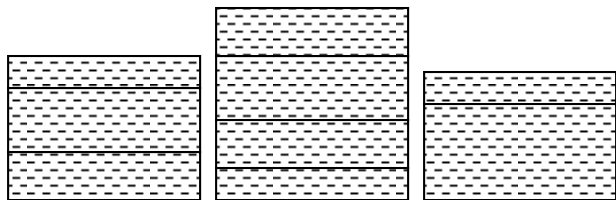
**C  
E  
T  
E  
M**  
**Público**





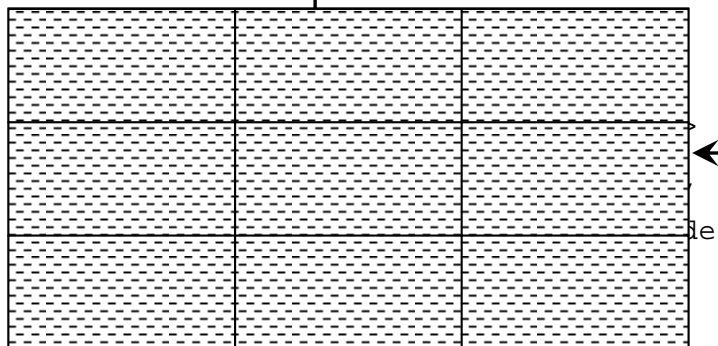
**C  
E  
T  
E  
M**

**Público**



	Quantity	By article	By paragraph
Articles	112.509	-	-
Extracts	489.033	4,35	-
Identified titles	269.848	0,42	0,55
Identified authors	47.763	0,42	0,10

(Texts from 1996 to 1998)

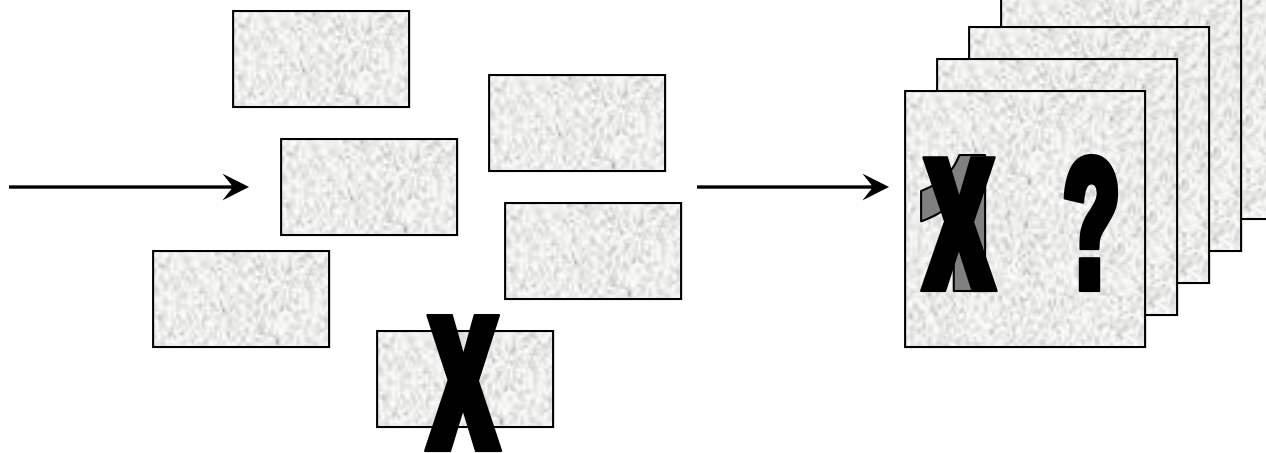
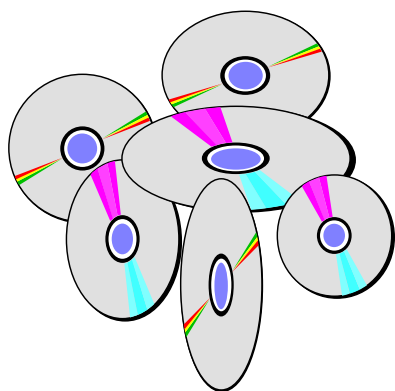


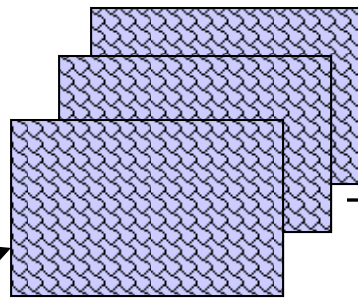
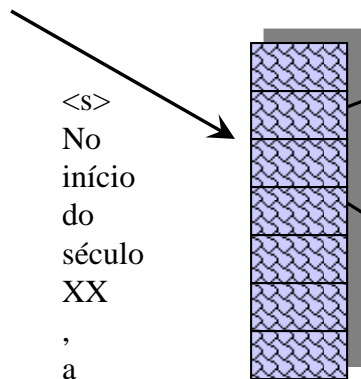
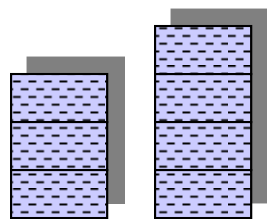
```

<t>A «farsa» ortográfica </t>
<a>João Bosco Jardim, em Brasília</a>
Em vez de uma folha em branco, o
Acordo Ortográfico será uma página de
palavras riscadas e emendadas .

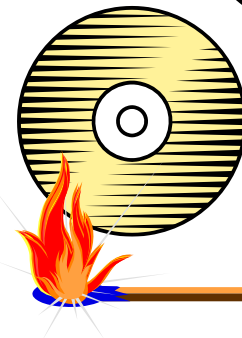
```

**C  
E  
T  
E  
M  
Público**





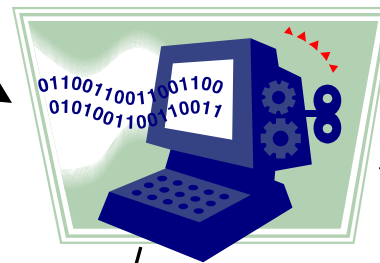
# Text



<s>  
No  
início  
do  
século  
XX  
,  
a  
localidade  
norueguesa  
de

93a Hammerfest SOC

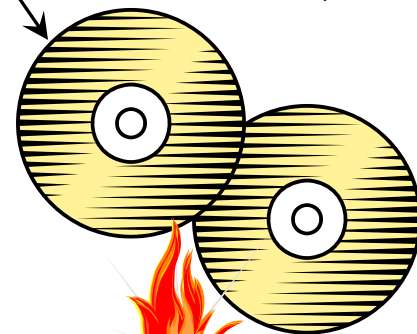
foi  
o  
primeiro  
local  
da  
Europa  
a  
ter  
luz  
eléctrica  
nas  
ruas  
.  
</s>



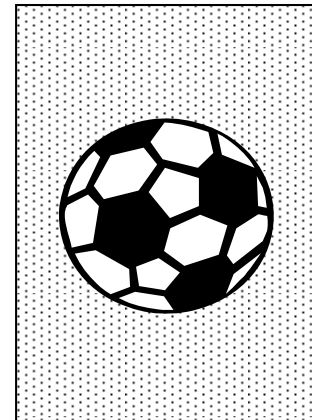
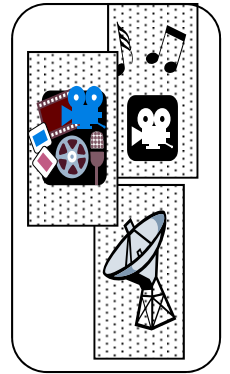
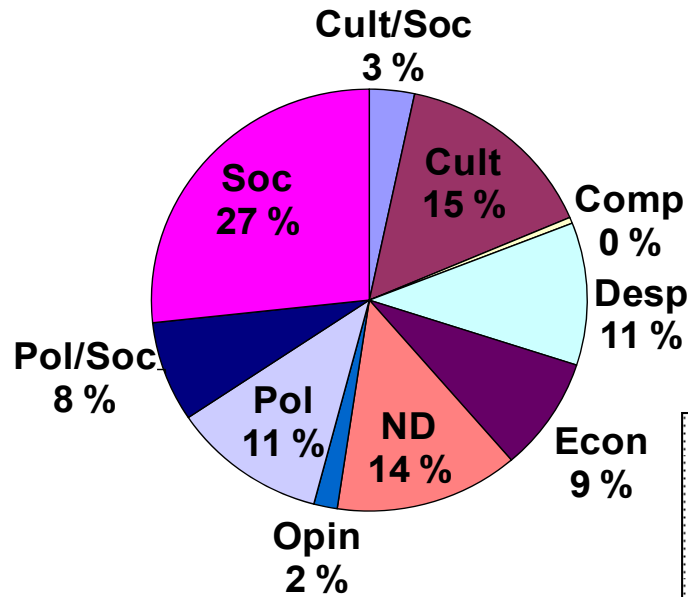
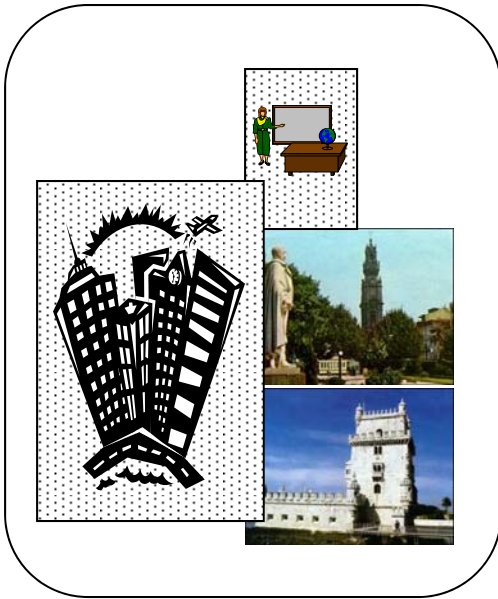
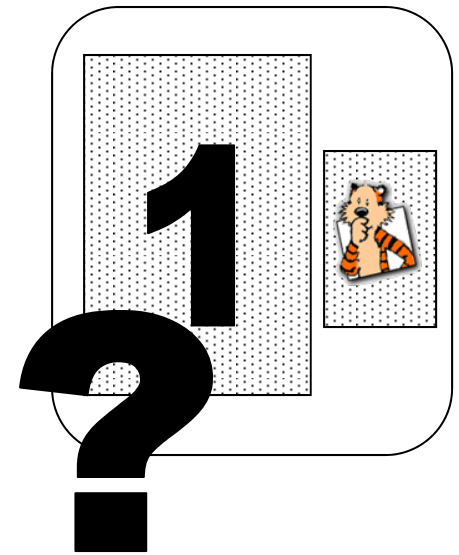
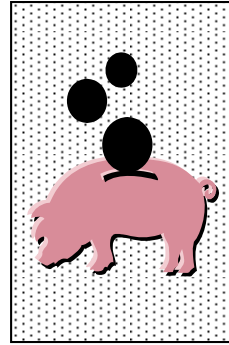
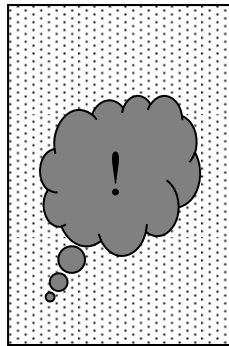
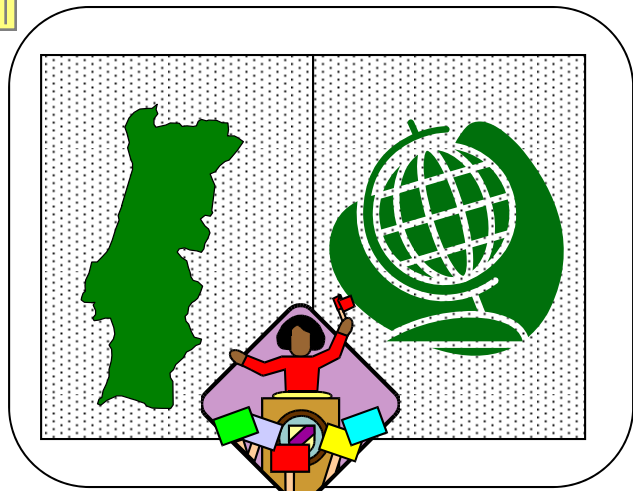
AC/DC



# CQP



# C E T E M Público

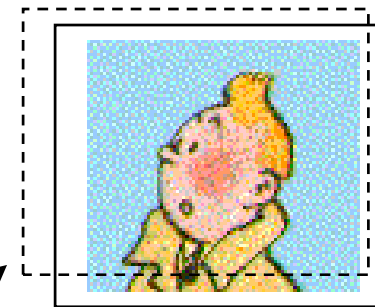


C  
E  
T  
E  
M

Público



- ✓ P.R.
- ✓ J.V.M.
- ✓ Mário Soares
- ✓ José Vítor Malheiros
- ✓ Fernando H. Cardoso
- ✓ Tolentino da Nóbrega
- ✓ Isabel Dias e Cunha
- ✓ Margarida Gomes e Vítor de Sousa
- ✓ Karim El-Ghawry, no Cairo
- C** ✓ do nosso enviado, Paulo Rocha
- E** ✗ Estados Unidos
- T** ✗ Faculdade de Ciências
- E** ✗ PERGUNTA DO DIA
- M** ✓ Vinho do Porto



Mal identificados: 4%

Não identificados: 3%  
(valores sobre uma amostra)

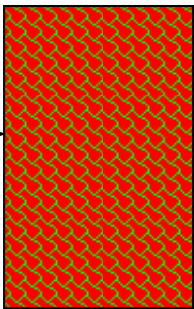
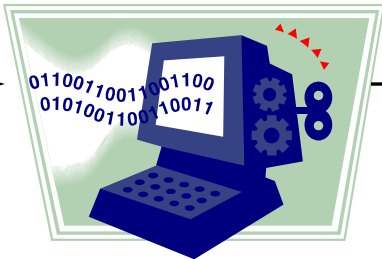
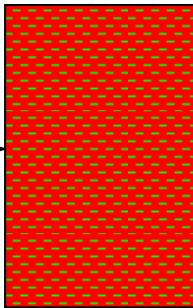
<t> Jorge Sampaio em Bucareste </t>

<a> Jorge Sampaio, em Bucareste </a>

# Público



*Jornal ?*  
*Edição on-line*



**C**  
**E**  
**T**  
**E**  
**M**

Space {



CETEMPúblico

Corpus Experimental

Time {



CETEMPúblico

Corpus Experimental

**Público**

**Resultados da procura**

Procura: "**Internet**" %c;  
Distribuição de **semestre**  
Corpus: CETEMPúblico 1.1

8883 casos.

**Distribuição**

98b	1146
98a	1931
97b	1334
97a	756
96b	409
96a	503
95b	1108
95a	1197
94b	279
94a	195
93b	3
93a	21
92a	1

Ext **332461** (nd, 96b): A oposição norueguesa criticara recentemente o ministro do Petróleo, **Jens Stoltenberg**, e o conjunto do Governo trabalhista, de não estarem suficientemente atentos à situação dos direitos humanos na Indonésia .

Ext **383637** (pol-soc, 94b): Ontem, o ministro da **Industria** e Energia, **Jens Stoltenberg**, foi obrigado a cancelar o vôo para Bruxelas porque o referendo lhe bloqueou o acesso à reunião dos ministros da Energia dos Doze .

Ext **582679** (soc, 95a): «Trata-se de encontrar um local seguro para atracar temporariamente, enquanto a Shell elabora um plano alternativo para a desactivação da Brent Spar», afirmou o ministro norueguês **Jens Stoltenberg** .

Ext **699917** (pol, 96b): Com Jagland sobe também o ministro da Indústria e Energia, **Jens Stoltenberg**, que transita para as Finanças, o que pode ser interpretado como um prémio para o seu bom desempenho na gestão dos (vastos) recursos petrolíferos do país -- que valem a este político muitos pontos nas sondagens .

Ext **1526329** (soc, 95a): Para já, no que respeita à Brent Spar, a Noruega declarou-se disposta a receber provisoriamente a plataforma, segundo declarações do ministro **Jens Stoltenberg**, que foi contactado pelo seu homólogo holandês, Hans Wijers .

**Resultados da procura**

Procura: "**Re.(k|que).av.\***".  
Pedido: Distribuição das formas  
Corpus: CETEMPúblico 1.1

123 casos.

**Distribuição**

Reiquejavique	55
Reykjavik	46
Reikjavik	10
Reikjavic	5
Reikjavyk	3
Reykjavic	3
Reikyavik	1

**Resultados da procura**

Procura: "**Rosenborg**".  
Distribuição de **assunto**  
Corpus: CETEMPúblico 1.1

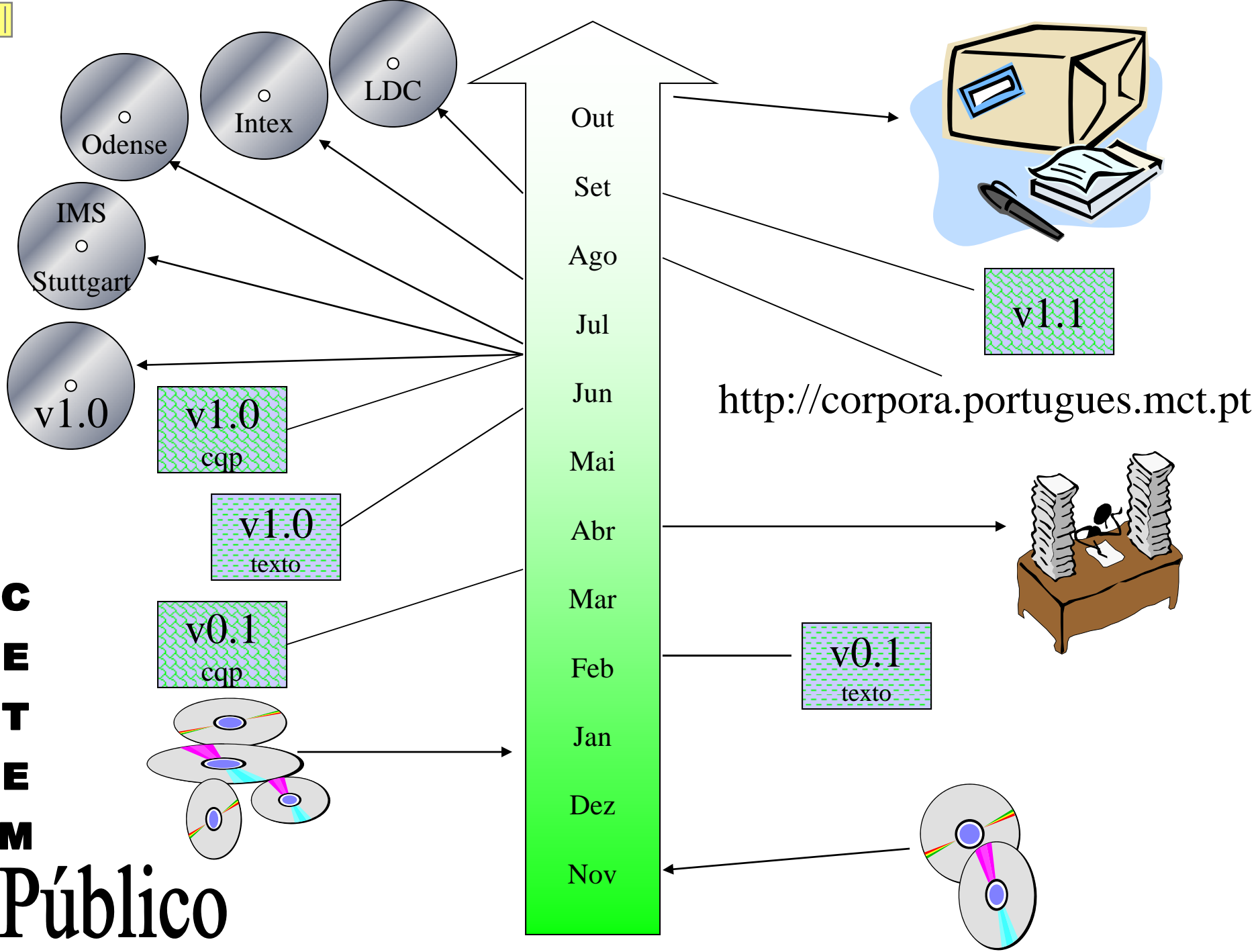
309 casos.

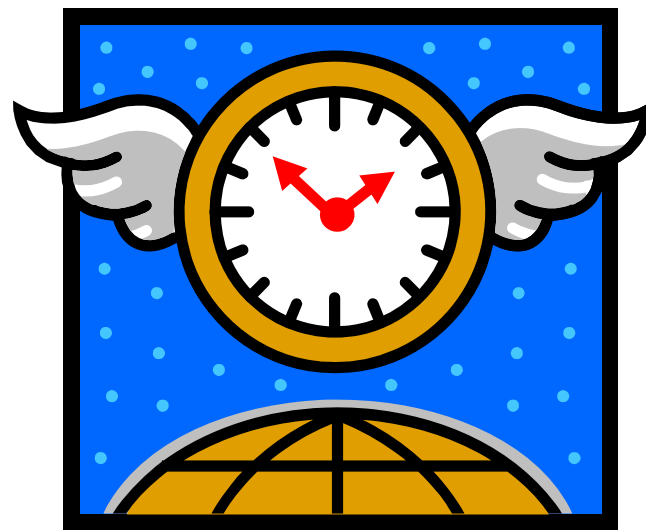
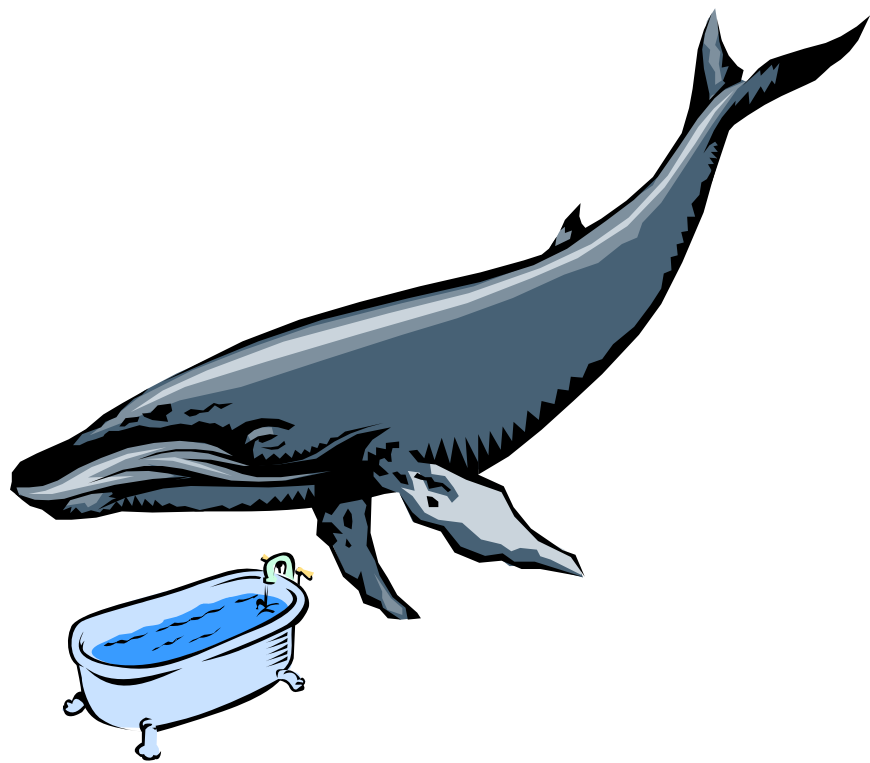
**Distribuição**

des	259
nd	42
opi	4
soc	3
clt	1



**C  
E  
T  
E  
M**  
**Público**

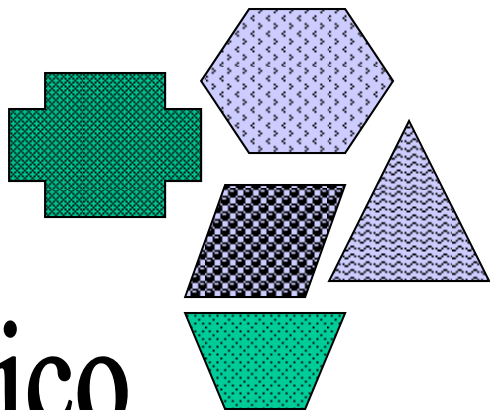




2 X

C  
E  
T  
E  
M

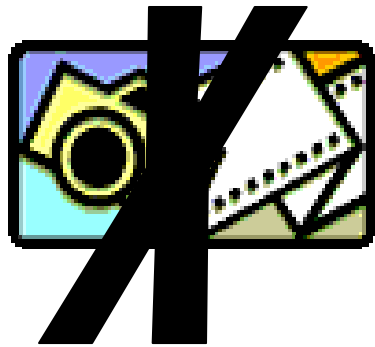
Público





v1.1

Rose	4
Brand	47
Vikin	45
Tro	44
St	42

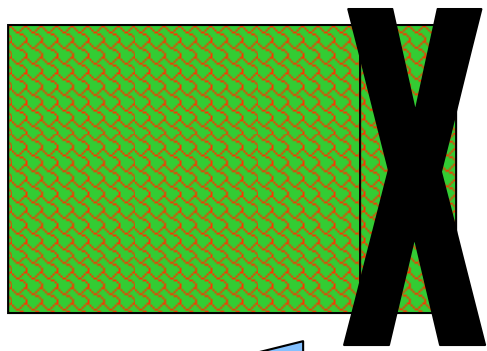


C  
E  
T  
E  
M

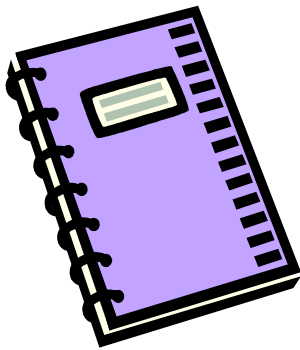
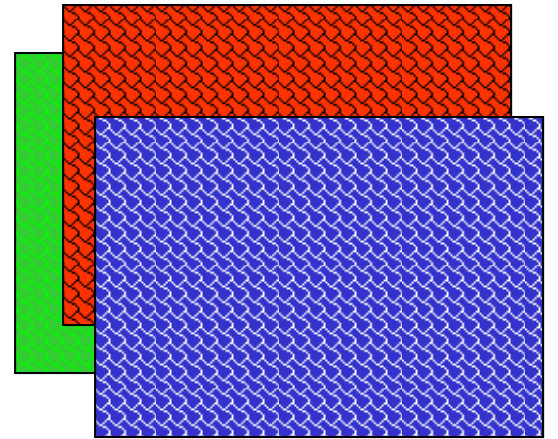
Público



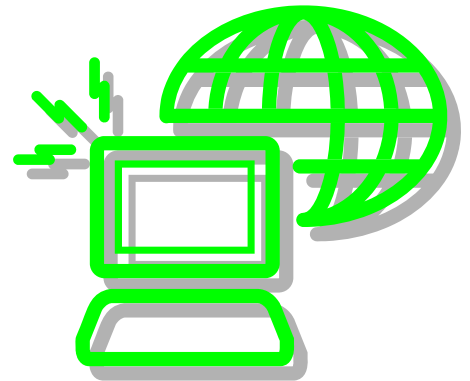
<ext sec=nd sem=94b>



Fu t t u r o



~~t~~ → <t>  
~~a~~ → <a>



C  
E  
T  
E  
M

Público



<http://www.portugues.mct.pt/>

<http://cgi.portugues.mct.pt/cetempublico>