


Text retrieval systems

Jon Bing
Norwegian Research Center for Computers and Law

NRCCCL (University of Oslo, Faculty of Law)



Analysing natural language documents


Article 3
Internal market
1. Each Member State shall ensure that the information society services provided by a service provider established on its territory comply with the national provisions applicable in the Member State in question which fall within the coordinated field

Identify

- Document number
- Paragraph number
- Sentence number
- Word number

Four tired address
(D/P/S/W)

NRCCCL (University of Oslo, Faculty of Law)



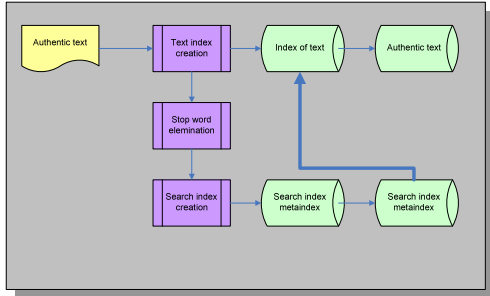
Terminators

- Eliminate word terminator:
. , " !) (
- Eliminate sentence terminators
. ! ?

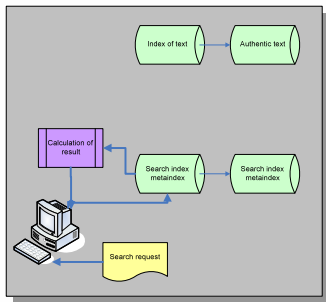
Define word, sentence, paragrah and document

NRCCCL (University of Oslo, Faculty of Law)

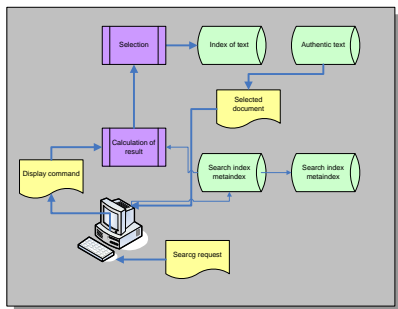
Text retrieval construction



Retrieval – I



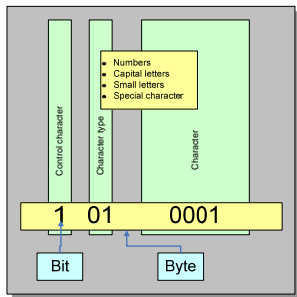
Retrieval - II



ASCII Character representation

Description	Char	HEX	ASCII
small a	a	61	a
small b	b	62	b
small c	c	63	c
small d	d	64	d
small e	e	65	e
small f	f	66	f

Character structure



Subtracting b from a

$$\begin{array}{r}
 1\ 01\ 00001 \\
 - 1\ 01\ 00010 \\
 \hline
 - 0\ 00\ 00001
 \end{array}$$

Search index

Article 3

Internal market

1. Each Member State shall ensure that the information society services provided by a service provider established on its territory comply with the national provisions applicable in the Member State in question which

Word	Doc no	Para no	Sentence no	Word no
1	1	3	1	1
3	1	1	1	2
Article	1	1	1	1
Each	1	3	2	1
ensure	1	3	2	5
Internal	1	2	1	1
Market	1	2	1	2
Member	1	3	2	2
shall	1	3	2	4
State	1	3	2	3
