

Università di Roma “La Sapienza”

**DIPLOMA DI STATISTICA
PROGETTO CAMPUS**

**Guida all’uso dello Spad-T
Applicazione “Texte”**

Luca Giuliano – Alessandro Stabellini

Anno 2000

DOS

Sigla della locuz. ingl. Disk Operating System «sistema operativo a disco», con cui si indica in informatica una classe di sistemi operativi in grado di gestire memorie di massa ad accesso diretto (dischi, tamburi, ecc.).

Il DOS è un sistema operativo:

- A riga di comando: i comandi vengono impartiti dal *prompt*
- Monoutente: non è possibile la condivisione delle risorse
- Monotask: non è possibile eseguire più processi alla volta

Il programma

Lo Spad-T

1. È un applicativo per [DOS](#) in lingua francese che può essere avviato anche in “finestra” sotto Windows

2. Ha un’*interfaccia utente* di tipo grafico (*shell*) che consente di impostare il lavoro evitando l’apprendimento della sintassi

3. È orientato all’Analisi Statistica di un testo secondo un approccio [lessicometrico](#)

4. Fornisce gli strumenti computazionali utili alla Statistica Testuale.

Lessicometria

Settore della lessicologia che applica i procedimenti statistici per lo studio e la descrizione del lessico in generale o del vocabolario di un autore.

Il programma

Lo Spad-T

<p>1. È un applicativo per <u>DOS</u> in lingua francese che può essere avviato anche in “finestra” sotto Windows</p>	<p>2. Ha un' <i>interfaccia utente</i> di tipo grafico (<i>shell</i>) che consente di impostare il lavoro evitando l'apprendimento della sintassi</p>
<p>3. È orientato all'Analisi Statistica di un testo secondo un approccio <u>lessicometrico</u></p>	<p>4. Fornisce gli strumenti computazionali utili alla Statistica Testuale.</p>

Gli strumenti dello Spad-T

- Numerizzazione parole: *NUMER*
- Segmenti ripetuti: *SEGME*
- Correzione del vocabolario: *CORTE*
- Riduzione soglia frequenza: *SETEX*
- Concordanza delle parole: *CORDA*
- Parole o segmenti caratteristici: *MOCAR*
- Risposte caratteristiche: *RECAR*
- An. Corrispondenze semplici: *APLUM*
- An. Corrispondenze multiple: *ASPAR*
- Creazione di *cluster*: *PARTI*
- Interpretazione dei fattori: *CLAIR*

Segmenti Ripetuti

Un segmento ripetuto è una sequenza di parole (*sintagma*), tra tutte le disposizioni di $2, 3, \dots, q$ parole che si ripetono più volte nel *corpus*, contenuta interamente in un *frammento* di testo.

Ad esempio una frase del tipo "il governo si propone di" contribuisce alle occorrenze di segmenti quali: <il governo>, <il governo si>, <il governo si propone>, <governo si>, <governo si propone>, <si propone>, <si propone di>, oltre che al segmento <il governo si propone di>.

Gli strumenti dello Spad-T

- Numerizzazione parole: *NUMER*
- Segmenti ripetuti: *SEGME*
- Correzione del vocabolario: *CORTE*
- Riduzione soglia frequenza: *SETEX*
- Concordanza delle parole: *CORDA*
- Parole o segmenti caratteristici: *MOCAR*
- Risposte caratteristiche: *RECAR*
- An. Corrispondenze semplici: *APLUM*
- An. Corrispondenze multiple: *ASPAR*
- Creazione di *cluster*: *PARTI*
- Interpretazione dei fattori: *CLAIR*

Concordanze

Le *concordanze* sono un elenco di brevi estratti di testo. Ogni estratto si chiama *concordanza*. Gli estratti sono fatti in modo da mostrare il testo che si trova prima e dopo una data parola.

Per esempio, le concordanze della parola *casa* sono gli estratti di testo che contengono la parola *casa*.

Gli strumenti dello Spad-T

- Numerizzazione parole: *NUMER*
- Segmenti ripetuti: *SEGME*
- Correzione del vocabolario: *CORTE*
- Riduzione soglia frequenza: *SETEX*
- Concordanza delle parole: *CORDA*
- Parole o segmenti caratteristici: *MOCAR*
- Risposte caratteristiche: *RECAR*
- An. Corrispondenze semplici: *APLUM*
- An. Corrispondenze multiple: *ASPAR*
- Creazione di *cluster*: *PARTI*
- Interpretazione dei fattori: *CLAIR*

Parole/Segmenti caratteristici

Una parola/segmento è tanto più specifica/o per un *testo* quanto più la sua frequenza totale nel *corpus* è assorbita dalla sua stessa frequenza all'interno del *testo*.

Alla base del calcolo delle specificità vi è un articolato test di natura statistica.

Gli strumenti dello Spad-T

- Numerizzazione parole: *NUMER*
- Segmenti ripetuti: *SEGME*
- Correzione del vocabolario: *CORTE*
- Riduzione soglia frequenza: *SETEX*
- Concordanza delle parole: *CORDA*
- Parole o segmenti caratteristici: *MOCAR*
- Risposte caratteristiche: *RECAR*
- An. Corrispondenze semplici: *APLUM*
- An. Corrispondenze multiple: *ASPAR*
- Creazione di *cluster*: *PARTI*
- Interpretazione dei fattori: *CLAIR*

Corrispondenze semplici

parole*testi

Forma		Testo			
		Testo1	Testo2	Testo3	Testo4
1	DA	82	35	40	145
2	PER	56	77	19	70
3	FORSE	49	62	33	12
...	
i	...	29	10	25	56
...	
$V_{(s)}$	SU	3	1	5	7

Gli strumenti dello Spad-T

- Numerizzazione parole: *NUMER*
- Segmenti ripetuti: *SEGME*
- Correzione del vocabolario: *CORTE*
- Riduzione soglia frequenza: *SETEX*
- Concordanza delle parole: *CORDA*
- Parole o segmenti caratteristici: *MOCAR*
- Risposte caratteristiche: *RECAR*
- An. Corrispondenze semplici: *APLUM*
- An. Corrispondenze multiple: *ASPAR*
- Creazione di *cluster*: *PARTI*
- Interpretazione dei fattori: *CLAIR*

Come lavora lo Spad-T

Lo Spad-T riconosce ogni parola di un testo come forma grafica , ovvero come un elemento il cui tratto caratteristico è l'apparenza grafica (parole tali e quali scritte nel testo: poiché diverso da poichè) e basa l'elaborazione:

sulla [tabella lessicale](#) intera riferita al *corpus* oggetto di studio;

sulla frequenza d'occorrenza delle parole del *corpus*;

sul confronto fra "profili" lessicali delle differenti parti del *corpus*.

Tabella lessicale

****TESTO1

Questo	è	il	primo	testo	di	prova	riga	uno
1	2	3	4	5	6	7	8	9
Questo	è	Il	primo	testo	di	prova	riga	due
1	2	3	4	5	6	7	8	10

****TESTO2

Questo	è	il	secondo	testo	di	prova	riga	uno
1	2	3	11	5	6	7	8	9
Questo	è	Il	secondo	testo	di	prova	riga	due
1	2	3	11	5	6	7	8	10

Classificazione delle parole per riga e testi in cui occorrono

Testo Forma	Testo1		Totale	Testo Forma	Testo2		Totale
	Riga1	Riga2			Riga1	Riga2	
1	1	1	2	1	1	1	2
2	1	1	2	2	1	1	2
3	1	1	2	3	1	1	2
4	1	1	2	4	0	0	0
5	1	1	2	5	1	1	2
6	1	1	2	6	1	1	2
7	1	1	2	7	1	1	2
8	1	1	2	8	1	1	2
9	1	0	1	9	1	0	1
10	0	1	1	10	0	1	1
11	0	0	0	11	1	1	2
Totale	9	9	18	Totale	9	9	18

Come lavora lo Spad-T

Lo Spad-T riconosce ogni parola di un testo come forma grafica , ovvero come un elemento il cui tratto caratteristico è l'apparenza grafica (parole tali e quali scritte nel testo: poiché diverso da poichè) e basa l'elaborazione:

sulla [tabella lessicale](#) intera riferita al *corpus* oggetto di studio;

sulla frequenza d'occorrenza delle parole del *corpus*;

sul confronto fra "profili" lessicali delle differenti parti del *corpus*.

Il file di dati

Per costruire un file di dati testuale (*corpus*) è necessario:

- Un editor di testi
- Alcune regole di “sintassi”.

Editor di testi

È necessario un editor di testi per DOS, ma può essere utilizzato un editor di Windows.

In questo caso bisogna avere accortezza di:

- Salvare il testo in formato txt con interruzione di riga (lo Spad-T non legge testo oltre le 80 colonne);
- Leggere i risultati in Windows.

Il formato txt

Un file di testo con estensione txt creato con un editor per Windows userà una tabella codice ANSI Windows, diversa da quella utilizzata dal DOS.

Ciò significa che un carattere sotto DOS può avere una forma grafica totalmente diversa in Windows

Tabella codice

definizione

Una tabella codice è una tabella in cui vi è una corrispondenza biunivoca tra un carattere ed il corrispondente valore decimale o esadecimale. La tabella codice [ANSI Windows](#), utilizzata dalle applicazioni Windows è **DIVERSA** [da quella utilizzata dal DOS](#)

Tabella Codice

ANSI Windows

Tabella codice ANSI Windows® Codice pagina 850 Courier font

	0	1	2	3	4	5	6	7	8	9	A	B	C	D	E	F
00:	NUL	SOH	STX	ETX	EOT	ENQ	ACK	BEL	BS	HT	LF	VT	FF	CR	SO	SI
10:	DLE	DC1	DC2	DC3	DC4	NAK	SYN	ETB	CAN	EM	SUB	ESC	FS	GS	RS	US
20:		!	"	#	\$	%	&	'	()	*	+	,	-	.	/
30:	0	1	2	3	4	5	6	7	8	9	:	;	<	=	>	?
40:	@	A	B	C	D	E	F	G	H	I	J	K	L	M	N	O
50:	P	Q	R	S	T	U	V	W	X	Y	Z	[\]	^	_
60:	`	a	b	c	d	e	f	g	h	i	j	k	l	m	n	o
70:	p	q	r	s	t	u	v	w	x	y	z	{		}	~	DEL
80:	€	□	/	f	~	...	†	‡	^	‰	Š	<	œ	□	□	□
90:	□	`	/	ˆ	˜	•	-	—	˘	™	š	>	œ	□	□	ÿ
A0:		¡	¢	£	¤	¥	¦	§	¨	©	ª	«	¬	­	®	¯
B0:	°	±	²	³	´	µ	¶	·	¸	¹	º	»	¼	½	¾	¿
C0:	À	Á	Â	Ã	Ä	Å	Æ	Ç	È	É	Ê	Ë	Ì	Í	Î	Ï
D0:	Ð	Ñ	Ò	Ó	Ô	Õ	Ö	×	Ø	Ù	Ú	Û	Ü	Ý	Þ	ß
E0:	à	á	â	ã	ä	å	æ	ç	è	é	ê	ë	ì	í	î	ï
F0:	ð	ñ	ò	ó	ô	õ	ö	÷	ø	ù	ú	û	ü	ý	þ	ÿ

Tabella codice

definizione

Una tabella codice è una tabella in cui vi è una corrispondenza biunivoca tra un carattere ed il corrispondente valore decimale o esadecimale. La tabella codice [ANSI Windows](#), utilizzata dalle applicazioni Windows è **DIVERSA** [da quella utilizzata dal DOS](#)

Tabella Codice

Microsoft DOS

Tabella codice Microsoft Dos® Codice pagina 850 Courier font

	0	1	2	3	4	5	6	7	8	9	A	B	C	D	E	F
00:	NUL	SOH	STX	ETX	EOT	ENQ	ACK	BEL	BS	HT	LF	VT	FF	CR	SO	SI
10:	DLE	DC1	DC2	DC3	DC4	NAK	SYN	ETB	CAN	EM	SUB	ESC	FS	GS	RS	US
20:		!	"	#	\$	%	&	'	()	*	+	,	-	.	/
30:	0	1	2	3	4	5	6	7	8	9	:	;	<	=	>	?
40:	@	A	B	C	D	E	F	G	H	I	J	K	L	M	N	O
50:	P	Q	R	S	T	U	V	W	X	Y	Z	[\]	^	_
60:	`	a	b	c	d	e	f	g	h	i	j	k	l	m	n	o
70:	p	q	r	s	t	u	v	w	x	y	z	{		}	~	DEL
80:	Ç	ù	é	â	ä	à	ã	ç	ê	ë	è	ì	í	î	ï	Ä
90:	É	æ	Æ	ô	ö	ò	û	ü	ÿ	Ö	Ü	ø	£	©	*	f
AO:	á	í	ó	ú	ñ	Ñ	ª	º	¿	@	¬	½	¼	;	«	»
BO:	::	::	■			Á	Â	À	©	¶	¶	¶	¶	;	¥	
CO:	L	L	T		-		ä	ã	L	¶	¶	¶	¶	=	¶	¶
DO:	ø	ð	Ê	Ë	È	ı	í	î	ï	ı	ı	ı	ı	ı	ı	ı
EO:	Ó	B	Ô	Ò	Ö	Õ	µ	þ	ƒ	Ú	Û	Ü	ý	ÿ	-	'
FO:	-	±	=	‰	¶	§	÷	,	°	˙	.	ı	ı	ı	ı	ı

Tabella codice

definizione

Una tabella codice è una tabella in cui vi è una corrispondenza biunivoca tra un carattere ed il corrispondente valore decimale o esadecimale. La tabella codice [ANSI Windows](#), utilizzata dalle applicazioni Windows è **DIVERSA** [da quella utilizzata dal DOS](#)

Il formato txt

Un file di testo con estensione txt creato con un editor per Windows userà una tabella codice ANSI Windows, diversa da quella utilizzata dal DOS.

Ciò significa che un carattere sotto DOS può avere una forma grafica totalmente diversa in Windows

I file in Windows e DOS

formato testo

File scritto in DOS letto in Windows

```
****Testo1
Questa § la prima riga del testo uno
Questa § la seconda riga del testo uno
****Testo2
Questa § la prima riga del testo due
Questa § la seconda riga el testo due
```

File scritto in Windows letto in Windows

```
****Testo1
Questa è la prima riga del testo uno
Questa è la seconda riga del testo uno
****Testo2
Questa è la prima riga del testo due
Questa è la seconda riga el testo due
```

[Formato binario](#)

I file in Windows e DOS

formato binario

File scritto in DOS letto in Windows

0:	2A 2A 2A 2A 54 65 73 74	6F 31 0D 0A 51 75 65 73	****Testo1..Ques
10:	74 61 20 8A 20 6C 61 20	70 72 69 6D 61 20 72 69	ta Š la prima ri
20:	67 61 20 64 65 6C 20 74	65 73 74 6F 20 75 6E 6F	ga del testo uno
30:	0D 0A 51 75 65 73 74 61	20 8A 20 6C 61 20 73 65	..Questa Š la se
40:	63 6F 6E 64 61 20 72 69	67 61 20 64 65 6C 20 74	conda riga del t
50:	65 73 74 6F 20 75 6E 6F	0D 0A 2A 2A 2A 2A 54 65	esto uno..****Te
60:	73 74 6F 32 0D 0A 51 75	65 73 74 61 20 8A 20 6C	sto2..Questa Š l
70:	61 20 70 72 69 6D 61 20	72 69 67 61 20 64 65 6C	a prima riga del
80:	20 74 65 73 74 6F 20 64	75 65 0D 0A 51 75 65 73	testo due..Ques
90:	74 61 20 8A 20 6C 61 20	73 65 63 6F 6E 64 61 20	ta Š la seconda
A0:	72 69 67 61 20 65 6C 20	74 65 73 74 6F 20 64 75	riga el testo du
B0:	65		e

[Tabella DOS](#)

File scritto in Windows letto in Windows

0:	2A 2A 2A 2A 54 65 73 74	6F 31 0D 0A 51 75 65 73	****Testo1..Ques
10:	74 61 20 E8 20 6C 61 20	70 72 69 6D 61 20 72 69	ta è la prima ri
20:	67 61 20 64 65 6C 20 74	65 73 74 6F 20 75 6E 6F	ga del testo uno
30:	0D 0A 51 75 65 73 74 61	20 E8 20 6C 61 20 73 65	..Questa è la se
40:	63 6F 6E 64 61 20 72 69	67 61 20 64 65 6C 20 74	conda riga del t
50:	65 73 74 6F 20 75 6E 6F	0D 0A 2A 2A 2A 2A 54 65	esto uno..****Te
60:	73 74 6F 32 0D 0A 51 75	65 73 74 61 20 E8 20 6C	sto2..Questa è l
70:	61 20 70 72 69 6D 61 20	72 69 67 61 20 64 65 6C	a prima riga del
80:	20 74 65 73 74 6F 20 64	75 65 0D 0A 51 75 65 73	testo due..Ques
90:	74 61 20 E8 20 6C 61 20	73 65 63 6F 6E 64 61 20	ta è la seconda
A0:	72 69 67 61 20 65 6C 20	74 65 73 74 6F 20 64 75	riga el testo du
B0:	65		e

[Tabella ANSI](#)

Tabella Codice

Microsoft DOS

Tabella codice Microsoft Dos® Codice pagina 850 Courier font

	0	1	2	3	4	5	6	7	8	9	A	B	C	D	E	F
00:	NUL	SOH	STX	ETX	EOT	ENQ	ACK	BEL	BS	HT	LF	VT	FF	CR	SO	SI
10:	DLE	DC1	DC2	DC3	DC4	NAK	SYN	ETB	CAN	EM	SUB	ESC	FS	GS	RS	US
20:		!	"	#	\$	%	&	'	()	*	+	,	-	.	/
30:	0	1	2	3	4	5	6	7	8	9	:	;	<	=	>	?
40:	@	A	B	C	D	E	F	G	H	I	J	K	L	M	N	O
50:	P	Q	R	S	T	U	V	W	X	Y	Z	[\]	^	_
60:	`	a	b	c	d	e	f	g	h	i	j	k	l	m	n	o
70:	p	q	r	s	t	u	v	w	x	y	z	{		}	~	DEL
80:	Ç	ù	é	â	ä	à	ã	ç	ê	ë	è	ì	í	ì	ä	Å
90:	É	æ	Æ	ô	ö	ò	û	ù	ÿ	Ö	Ü	ø	£	Ø	*	f
AO:	á	í	ó	ú	ñ	Ñ	ª	º	¿	@	¬	½	¼	;	«	»
BO:	::	::	■			Á	Â	À	©	¶		¶	¶	;	¥	
CO:	L	L	T		-		ä	ã	LL	¶		¶	¶	=	¶	
DO:	ø	Ð	Ê	Ë	È	ı	í	î	ï	┘	┘	■	■	-	ı	■
EO:	Ó	ß	Ô	Ò	Ö	Õ	µ	þ	ƒ	Ú	Û	Ü	ý	ÿ	-	'
FO:	-	±	=	‰	¶	§	÷	,	°	˙	.	ı	ı	ı	■	

I file in Windows e DOS

formato binario

File scritto in DOS letto in Windows

0:	2A 2A 2A 2A 54 65 73 74	6F 31 0D 0A 51 75 65 73	****Testo1..Ques
10:	74 61 20 8A 20 6C 61 20	70 72 69 6D 61 20 72 69	ta Š la prima ri
20:	67 61 20 64 65 6C 20 74	65 73 74 6F 20 75 6E 6F	ga del testo uno
30:	0D 0A 51 75 65 73 74 61	20 8A 20 6C 61 20 73 65	..Questa Š la se
40:	63 6F 6E 64 61 20 72 69	67 61 20 64 65 6C 20 74	conda riga del t
50:	65 73 74 6F 20 75 6E 6F	0D 0A 2A 2A 2A 2A 54 65	esto uno..****Te
60:	73 74 6F 32 0D 0A 51 75	65 73 74 61 20 8A 20 6C	sto2..Questa Š l
70:	61 20 70 72 69 6D 61 20	72 69 67 61 20 64 65 6C	a prima riga del
80:	20 74 65 73 74 6F 20 64	75 65 0D 0A 51 75 65 73	testo due..Ques
90:	74 61 20 8A 20 6C 61 20	73 65 63 6F 6E 64 61 20	ta Š la seconda
A0:	72 69 67 61 20 65 6C 20	74 65 73 74 6F 20 64 75	riga el testo du
B0:	65		e

[Tabella DOS](#)

File scritto in Windows letto in Windows

0:	2A 2A 2A 2A 54 65 73 74	6F 31 0D 0A 51 75 65 73	****Testo1..Ques
10:	74 61 20 E8 20 6C 61 20	70 72 69 6D 61 20 72 69	ta è la prima ri
20:	67 61 20 64 65 6C 20 74	65 73 74 6F 20 75 6E 6F	ga del testo uno
30:	0D 0A 51 75 65 73 74 61	20 E8 20 6C 61 20 73 65	..Questa è la se
40:	63 6F 6E 64 61 20 72 69	67 61 20 64 65 6C 20 74	conda riga del t
50:	65 73 74 6F 20 75 6E 6F	0D 0A 2A 2A 2A 2A 54 65	esto uno..****Te
60:	73 74 6F 32 0D 0A 51 75	65 73 74 61 20 E8 20 6C	sto2..Questa è l
70:	61 20 70 72 69 6D 61 20	72 69 67 61 20 64 65 6C	a prima riga del
80:	20 74 65 73 74 6F 20 64	75 65 0D 0A 51 75 65 73	testo due..Ques
90:	74 61 20 E8 20 6C 61 20	73 65 63 6F 6E 64 61 20	ta è la seconda
A0:	72 69 67 61 20 65 6C 20	74 65 73 74 6F 20 64 75	riga el testo du
B0:	65		e

[Tabella ANSI](#)

Tabella Codice

ANSI Windows

Tabella codice ANSI Windows® Codice pagina 850 Courier font

	0	1	2	3	4	5	6	7	8	9	A	B	C	D	E	F
00:	NUL	SOH	STX	ETX	EOT	ENQ	ACK	BEL	BS	HT	LF	VT	FF	CR	SO	SI
10:	DLE	DC1	DC2	DC3	DC4	NAK	SYN	ETB	CAN	EM	SUB	ESC	FS	GS	RS	US
20:		!	"	#	\$	%	&	'	()	*	+	,	-	.	/
30:	0	1	2	3	4	5	6	7	8	9	:	;	<	=	>	?
40:	@	A	B	C	D	E	F	G	H	I	J	K	L	M	N	O
50:	P	Q	R	S	T	U	V	W	X	Y	Z	[\]	^	_
60:	`	a	b	c	d	e	f	g	h	i	j	k	l	m	n	o
70:	p	q	r	s	t	u	v	w	x	y	z	{		}	~	DEL
80:	€	□	/	f	~	...	†	‡	^	‰	Š	<	œ	□	□	□
90:	□	`	/	ˆ	˜	•	-	—	˘	™	š	>	œ	□	□	ÿ
A0:		¡	¢	£	¤	¥	¦	§	¨	©	ª	«	¬	­	®	¯
B0:	°	±	²	³	´	µ	¶	·	¸	¹	º	»	¼	½	¾	¿
C0:	À	Á	Â	Ã	Ä	Å	Æ	Ç	È	É	Ê	Ë	Ì	Í	Î	Ï
D0:	Ð	Ñ	Ò	Ó	Ô	Õ	Ö	×	Ø	Ù	Ú	Û	Ü	Ý	Þ	ß
E0:	à	á	â	ã	ä	å	æ	ç	è	é	ê	ë	ì	í	î	ï
F0:	ð	ñ	ò	ó	ô	õ	ö	÷	ø	ù	ú	û	ü	ý	þ	ÿ

I file in Windows e DOS

formato binario

File scritto in DOS letto in Windows

0:	2A 2A 2A 2A 54 65 73 74	6F 31 0D 0A 51 75 65 73	****Testo1..Ques
10:	74 61 20 8A 20 6C 61 20	70 72 69 6D 61 20 72 69	ta Š la prima ri
20:	67 61 20 64 65 6C 20 74	65 73 74 6F 20 75 6E 6F	ga del testo uno
30:	0D 0A 51 75 65 73 74 61	20 8A 20 6C 61 20 73 65	..Questa Š la se
40:	63 6F 6E 64 61 20 72 69	67 61 20 64 65 6C 20 74	conda riga del t
50:	65 73 74 6F 20 75 6E 6F	0D 0A 2A 2A 2A 2A 54 65	esto uno..****Te
60:	73 74 6F 32 0D 0A 51 75	65 73 74 61 20 8A 20 6C	sto2..Questa Š l
70:	61 20 70 72 69 6D 61 20	72 69 67 61 20 64 65 6C	a prima riga del
80:	20 74 65 73 74 6F 20 64	75 65 0D 0A 51 75 65 73	testo due..Ques
90:	74 61 20 8A 20 6C 61 20	73 65 63 6F 6E 64 61 20	ta Š la seconda
A0:	72 69 67 61 20 65 6C 20	74 65 73 74 6F 20 64 75	riga el testo du
B0:	65		e

[Tabella DOS](#)

File scritto in Windows letto in Windows

0:	2A 2A 2A 2A 54 65 73 74	6F 31 0D 0A 51 75 65 73	****Testo1..Ques
10:	74 61 20 E8 20 6C 61 20	70 72 69 6D 61 20 72 69	ta è la prima ri
20:	67 61 20 64 65 6C 20 74	65 73 74 6F 20 75 6E 6F	ga del testo uno
30:	0D 0A 51 75 65 73 74 61	20 E8 20 6C 61 20 73 65	..Questa è la se
40:	63 6F 6E 64 61 20 72 69	67 61 20 64 65 6C 20 74	conda riga del t
50:	65 73 74 6F 20 75 6E 6F	0D 0A 2A 2A 2A 2A 54 65	esto uno..****Te
60:	73 74 6F 32 0D 0A 51 75	65 73 74 61 20 E8 20 6C	sto2..Questa è l
70:	61 20 70 72 69 6D 61 20	72 69 67 61 20 64 65 6C	a prima riga del
80:	20 74 65 73 74 6F 20 64	75 65 0D 0A 51 75 65 73	testo due..Ques
90:	74 61 20 E8 20 6C 61 20	73 65 63 6F 6E 64 61 20	ta è la seconda
A0:	72 69 67 61 20 65 6C 20	74 65 73 74 6F 20 64 75	riga el testo du
B0:	65		e

[Tabella ANSI](#)

Il file di dati

Per costruire un file di dati testuale (*corpus*) è necessario:

- Un editor di testi
- Alcune regole di “sintassi”.

La sintassi

Riferendoci all'applicazione "Texte" (o testo unico) dello Spad-T[®] utilizzata nel caso in cui non si abbia esigenza di associare al file di dati un file numerico per la codifica delle variabili

Avremo che

- ciascun testo, in cui è ripartito il *corpus* dovrà essere inserito tra due righe di testo che inizino con 4 asterischi più la eventuale etichetta identificativa del testo.
- Il *corpus* deve terminare con una riga composta da una sequenza di 4 segni di uguale (====).

Esempi di file dati

Applicazione "Texte"

File di dati: Un riga un individuo. (Per individuo si intende l'entità nella quale viene ripartito, a sua volta, il testo)

```

****Testo 1
Questa è la prima riga del testo uno -> Individuo 1_Testo 1
Questa è la seconda riga del testo uno -> Individuo 2_Testo 1
...
Questa è la i-esima riga del testo uno -> Individuo i_Testo 1
...
Questa e la n-esima riga del testo uno -> Individuo n_Testo 1
****Testo 2
Questa è la prima riga del testo due -> Individuo 1_Testo 2
Questa è la seconda riga del testo due -> Individuo 2_Testo 2
...
Questa è la i-esima riga del testo due -> Individuo i_Testo 2
...
Questa e la n-esima riga del testo due -> Individuo n_Testo 2
...
****Testo n
Questa è la prima riga del testo n -> Individuo 1_Testo n
Questa è la seconda riga del testo n -> Individuo 2_Testo n
...
Questa è la i-esima riga del testo n -> Individuo i_Testo n
...
Questa e la n-esima riga del testo n -> Individuo n_Testo n
=====
    
```

File di dati: Un individuo su più righe. (Per individuo si intende l'entità nella quale viene ripartito, a sua volta, il testo)

```

****Testo 1
Questa è la prima riga del testo uno } Individuo 1_Testo 1
Questa è la seconda riga del testo uno }
----
Questa è la terza riga del testo uno } Individuo 2_Testo 1
----
.....
Questa è la i-esima riga del testo uno } Individuo 3_Testo 1
...
Questa e la n-esima riga del testo uno }
****Testo 2
Questa è la prima riga del testo due } Individuo 1_Testo 2
Questa è la seconda riga del testo due }
...
Questa è la i-esima riga del testo due }
----
.....
Questa e la n-esima riga del testo due } Individuo n_Testo 2
...
****Testo n
Questa è la prima riga del testo n } Individuo 1_Testo n
----
Questa è la seconda riga del testo n }
...
Questa è la i-esima riga del testo n } Individuo 2_Testo n
...
Questa e la n-esima riga del testo n }
=====
    
```

Spad-T

Schermata Principale

Menù



Spad-T

Applicazione "Texte"

The image shows two overlapping screenshots of the Spad-T application interface. The top screenshot displays a menu with options: Fichiers, Commandes, Exécuter, éditer, and Graphiques. A sub-menu is open, showing 'nom du fichier TEXTE', 'Données numériques', and 'Créer application SPADT'. A red box highlights the 'Type de fichier' options: '1:Texte' and '2:Questions ouvertes'. The bottom screenshot shows the 'Créer une application SPAD.T texte' screen with various configuration parameters.

Menu Options: Fichiers, Commandes, Exécuter, éditer, Graphiques

Sub-menu Options: nom du fichier TEXTE, Données numériques, Créer application SPADT

Type de fichier Options: 1:Texte, 2:Questions ouvertes

Configuration Parameters:

- Un individu = Une ligne <oui:1 non:0>:1
- SELECTION DES MOTS
- Séparateurs faibles (sép. mots): , ' ; . : ' "
- Séparateurs forts (sép. de segments) ; . : < > . : < > ! ? ,
- Fréquence minimale des mots : 0
- Nombre maximum de mots retenus : 5000
- Longueur maximum d'un mot : 16
- Coefficient de taille du vocabulaire : 1
- Réservation de mémoire centrale MEM : 100000

Footer: CORPUS LAT, F1 Aide ; ESC annule champ ; F9 Sortir sans valider ; F10 valider