



JavaScript

String & JSON

自我介绍



王立 程序员
Nick Wang

百变空间CTO\联合创始人\董事、连续创业者
曾经在微软、阿里工作十年以上

18项发明专利

多家上市公司、基金顾问



Agenda

✓ String对象

✓ 构造函数

✓ 成员函数

✓ 字符集

✓ Unicode系列

✓ GBK系列

✓ JSON

✓ 动手实验 X 9

✓ 知识点 X 4

AJAX

XML

| String对象

动手实验 10-1

- ✓ 构造 String 对象
- ✓ string原始类型
- ✓ 区别

String对象的成员函数

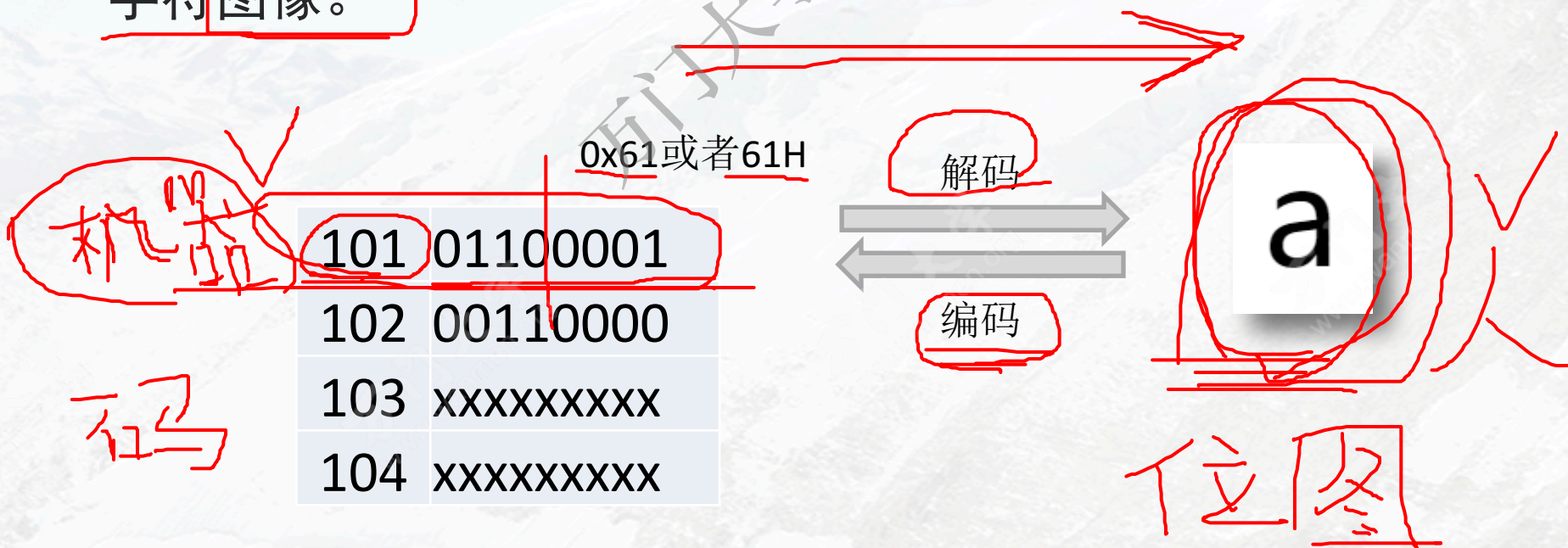
动手实验 10-2

- ✓ 数组方式访问字符串
- ✓ 字符串搜索：indexOf ()
- ✓ 字符串替换：replace ()
- ✓ 字符串转化维数组：split ()
- ✓ 特殊字符
- ✓ 大小写：toUpperCase () 、 toLowerCase ()
- ✓ 去除空格：trim () 、 trimLeft () 、 trimRight ()
- ✓ charAt () 、 charCodeAt () 、 fromCharCode ()

编码与解码

知识点 10-1

- ✓ 编码：用户输入字符，计算机记录对应二进制的值
- ✓ 解码：计算机根据记录的二进制值，显示成人类可以认识的字符图像。



字符集

知识点 10-2

- ✓ 字符集 (Charset) : 是一个系统支持的所有抽象字符的集合。
- ✓ 字符编码 (Character Encoding) : 字符集的编解码规则。
- ✓ 常见字符集:
 - ✓ ASCII字符集 0 127
 - ✓ GB字符集族:
 - ✓ GB2312、GBK、GB18030
 - ✓ Unicode字符集族:
 - ✓ UTF-8、UTF-16、UTF-32
 - ✓ UCS字符集族:
 - ✓ UCS-2、UCS-4

ASCII字符集

知识点 10-3

- ✓ ASCII (American Standard Code for Information Interchange, 美国标准信息交换码) :

- ✓ 128个
- ✓ 每个字符1字节
- ✓ 缺点：太小

Dec	Hex	Oct	Chr	Dec	Hex	Oct	HTML	Chr	Dec	Hex	Oct	HTML	Chr	Dec	Hex	Oct	HTML	Chr
0	0	000	NULL	32	20	040	 	Space	64	40	100	@	@	96	60	140	`	`
1	1	001	Start of Header	33	21	041	!	!	65	41	101	A	A	97	61	141	a	a
2	2	002	Start of Text	34	22	042	"	"	66	42	102	B	B	98	62	142	b	b
3	3	003	End of Text	35	23	043	#	#	67	43	103	C	C	99	63	143	c	c
4	4	004	End of Transmission	36	24	044	$	\$	68	44	104	D	D	100	64	144	d	d
5	5	005	Enquiry	37	25	045	%	%	69	45	105	E	E	101	65	145	e	e
6	6	006	Acknowledgment	38	26	046	&	&	70	46	106	F	F	102	66	146	f	f
7	7	007	Bell	39	27	047	'	'	71	47	107	G	G	103	67	147	g	g
8	8	010	Backspace	40	28	050	((72	48	110	H	H	104	68	150	h	h
9	9	011	Horizontal Tab	41	29	051))	73	49	111	I	I	105	69	151	i	i
10	A	012	Line feed	42	2A	052	*	*	74	4A	112	J	J	106	6A	152	j	j
11	B	013	Vertical Tab	43	2B	053	+	+	75	4B	113	K	K	107	6B	153	k	k
12	C	014	Form feed	44	2C	054	,	,	76	4C	114	L	L	108	6C	154	l	l
13	D	015	Carriage return	45	2D	055	-	-	77	4D	115	M	M	109	6D	155	m	m
14	E	016	Shift Out	46	2E	056	.	.	78	4E	116	N	N	110	6E	156	n	n
15	F	017	Shift In	47	2F	057	/	/	79	4F	117	O	O	111	6F	157	o	o
16	10	020	Data Link Escape	48	30	060	0	0	80	50	120	P	P	112	70	160	p	p
17	11	021	Device Control 1	49	31	061	1	1	81	51	121	Q	Q	113	71	161	q	q
18	12	022	Device Control 2	50	32	062	2	2	82	52	122	R	R	114	72	162	r	r
19	13	023	Device Control 3	51	33	063	3	3	83	53	123	S	S	115	73	163	s	s
20	14	024	Device Control 4	52	34	064	4	4	84	54	124	T	T	116	74	164	t	t
21	15	025	Negative Ack.	53	35	065	5	5	85	55	125	U	U	117	75	165	u	u
22	16	026	Synchronous idle	54	36	066	6	6	86	56	126	V	V	118	76	166	v	v
23	17	027	End of Trans. Block	55	37	067	7	7	87	57	127	W	W	119	77	167	w	w
24	18	030	Cancel	56	38	070	8	8	88	58	130	X	X	120	78	170	x	x
25	19	031	End of Medium	57	39	071	9	9	89	59	131	Y	Y	121	79	171	y	y
26	1A	032	Substitute	58	3A	072	:	:	90	5A	132	Z	Z	122	7A	172	z	z
27	1B	033	Escape	59	3B	073	;	;	91	5B	133	[[123	7B	173	{	{
28	1C	034	File Separator	60	3C	074	<	<	92	5C	134	\	\	124	7C	174	|	
29	1D	035	Group Separator	61	3D	075	=	=	93	5D	135]]	125	7D	175	}	}
30	1E	036	Record Separator	62	3E	076	>	>	94	5E	136	^	^	126	7E	176	~	~
31	1F	037	Unit Separator	63	3F	077	?	?	95	5F	137	_	_	127	7F	177		Del

GB2312字符集

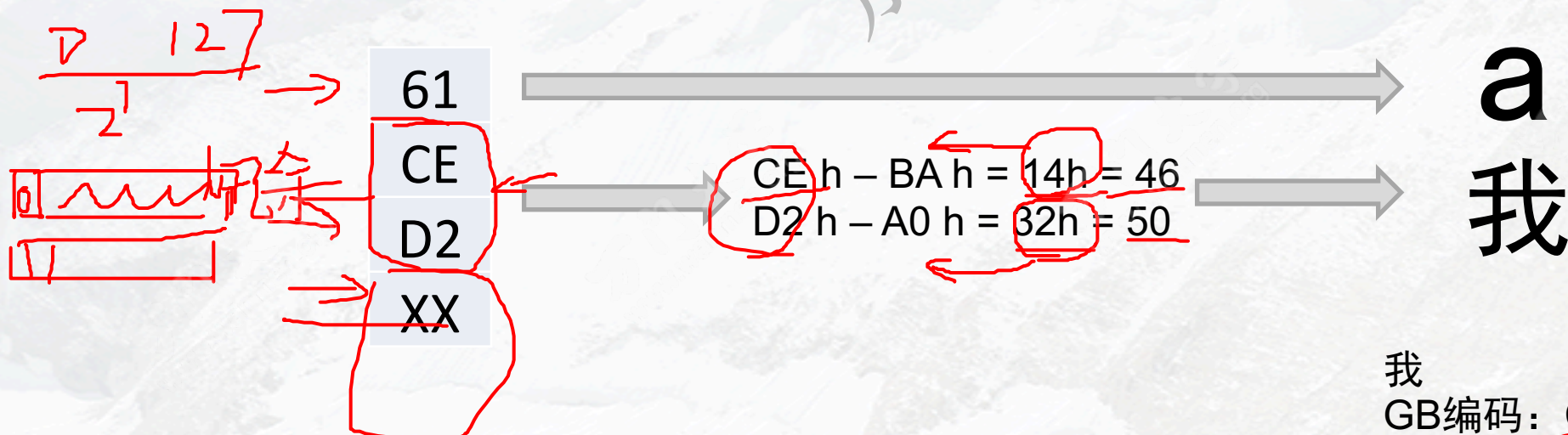
知识点 10-4

- ✓ 收录6763个汉字，以及682个字符
- ✓ ASCII字符使用1个字节，新收录的汉字字符用2个字节
- ✓ 兼容ASCII字符集
- ✓ 太小；变长字节数

老文件

新方法

方法



我
GB编码: CED2 h
区位码: 4650

GBK、GB18030字符集

知识点 10-4

✓ GBK字符集：

- ✓ 兼容GB2312、兼容Big5 → 兼容ASCII
- ✓ ASCII码1字节，GBK字符2字节
- ✓ 变长字节数。

✓ GB18030字符集：

- ✓ 兼容GBK → 兼容GB2312、兼容Big5 → 兼容ASCII
- ✓ ASCII码1字节，GB18030使用2字节或者4字节
- ✓ 变长字节数。

UTF-8字符集

知识点 10-5

ASCII

- ✓ Unicode值，世界上的任何字符都用2~4个字节表示。
- ✓ UTF-8, ASCII码用1字节表示，其他字符用2~6个字节表示，编码见下图。
- ✓ 兼容ASCII，不兼容GB系列

Unicode符号范围 | UTF-8编码方式
(十六进制) | (二进制)

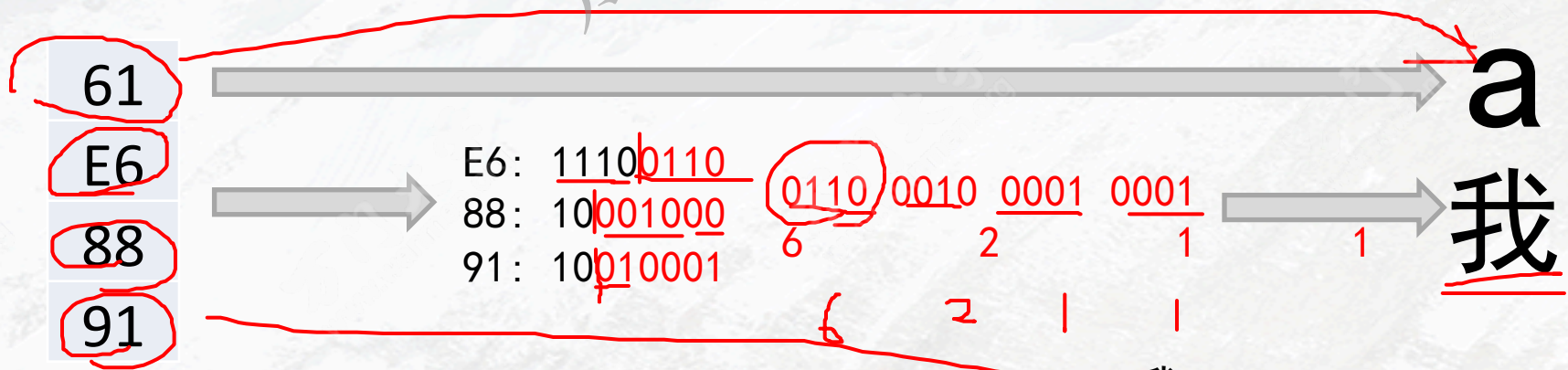
0000	0000-0000	007F		0xxxxxxx						
0000	0080-0000	07FF		110xxxxx	10xxxxxx	1				
0000	0800-0000	FFFF		1110xxxx	10xxxxxx	10xxxxxx	3			
0001	0000-001F	FFFF		11110xxx	10xxxxxx	10xxxxxx	10xxxxxx			
0020	0000-03FF	FFFF		11110xxx	10xxxxxx	10xxxxxx	10xxxxxx	10xxxxxx		
0400	0000-7FFF	FFFF		11110xxx	10xxxxxx	10xxxxxx	10xxxxxx	10xxxxxx	10xxxxxx	

UTF-8字符集

知识点 10-5

Unicode符号范围 | UTF-8编码方式

0000	0000-0000	007F		0xxxxxxx
0000	0080-0000	07FF		110xxxxx 10xxxxxx
0000	0800-0000	FFFF		1110xxxx 10xxxxxx 10xxxxxx
0001	0000-001F	FFFF		11110xxx 10xxxxxx 10xxxxxx 10xxxxxx
0020	0000-03FF	FFFF		11110xxx 10xxxxxx 10xxxxxx 10xxxxxx 10xxxxxx
0400	0000-7FFF	FFFF		11110xxx 10xxxxxx 10xxxxxx 10xxxxxx 10xxxxxx
				10xxxxxx

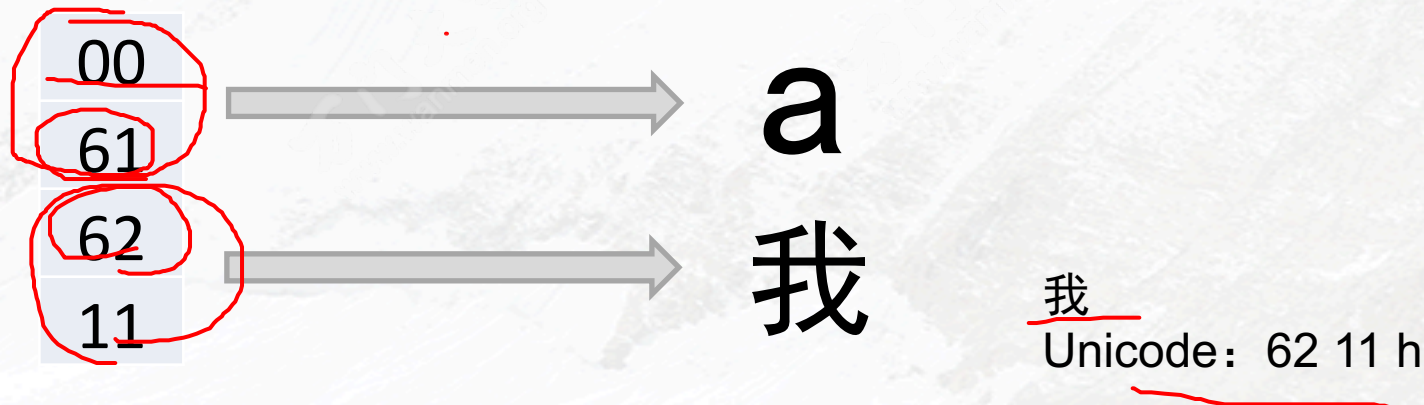


我
UTF-8编码: E6 88 91 h
Unicode: 62 11 h

UCS-2\UTF-16字符集

知识点 10-6

- ✓ UCS-2字符集，使用固定的2字节表示所有字符。
- ✓ 不兼容ASCII字符，每个ASCII字符在UCS-2中使用2字节表示。
- ✓ UTF-16，兼容UCS-2字符集，同时包含部分UCS-4字符集中的字符，UTF-16用2/4个字节表示字符。



UCS-4\UTF-32字符集

知识点 10-6

- ✓ UCS-4字符集，使用固定的4字节表示所有字符。
- ✓ 不兼容ASCII字符，每个ASCII字符在UCS-4中使用4字节表示，也不兼容UCS-2、UTF-16等。
- ✓ UCS-4，>兼容UTF-32字符集。

乱码问题

知识点 10-7

- ✓ 浏览器（JS引擎），通常遵循UCS-2字符集，支持部分UTF-16属性。
- ✓ HTML/CSS/JS文件字符集通常为UTF-8。
- ✓ 可能存在HTML/CSS/JS文件字符集为GBK的情况。
- ✓ 当文件的字符集和处理文件的程序（比如，浏览器，文本编辑器）字符集不相符时，程序解码后的文字不具有可读性，称为乱码。

HTTP协议

知识点 10-8

✓ HTTP Request & Response

✓ Resp: Content-Type

✓ Rqst: Accept-Charset

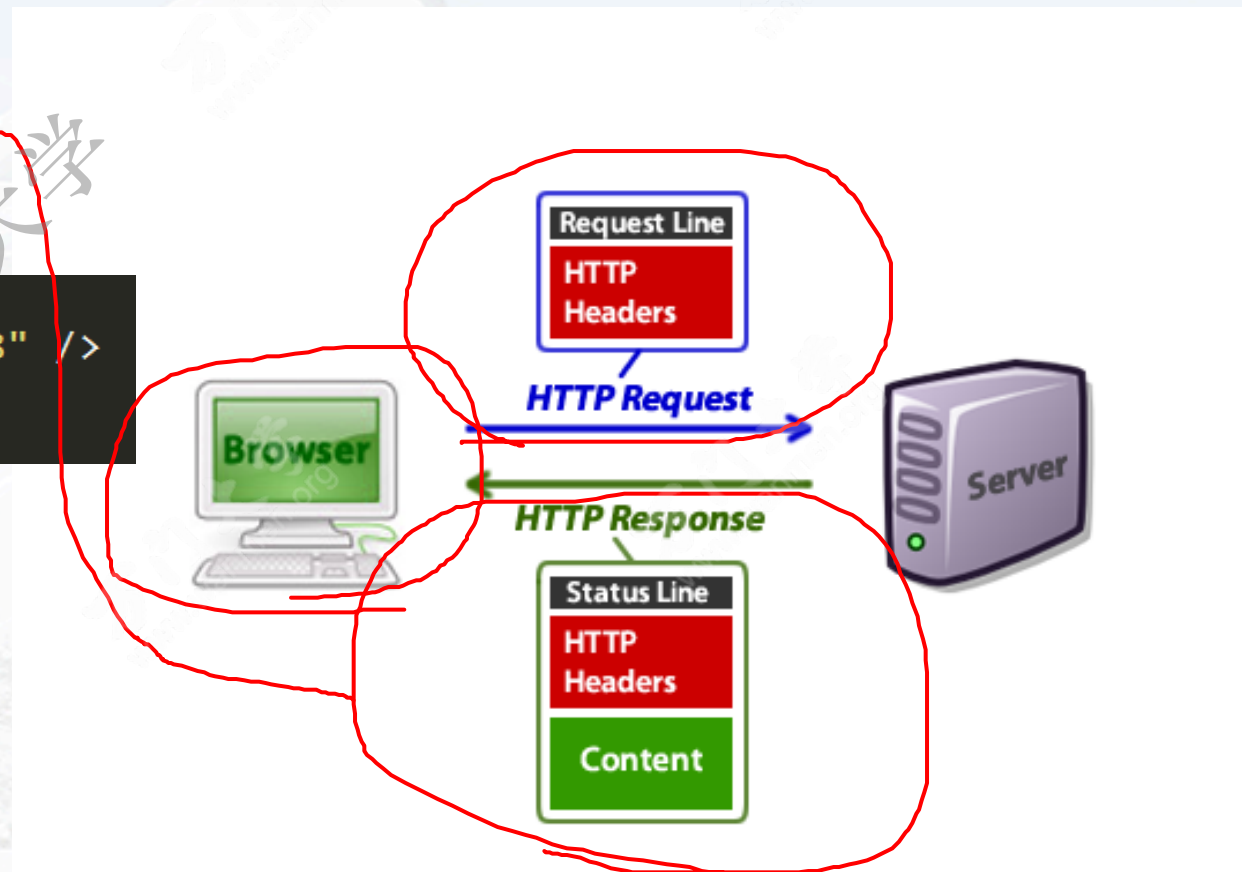
✓ HTML

✓ meta

```
<head>  
  <meta charset="utf-8" />  
  <title></title>  
</head>
```

✓ Content-Type优先级高于

meta



乱码问题

知识点 10-7

- ✓ 字符集：
 - ✓ 文本文件本身的字符集
 - ✓ 服务器通知浏览器使用何种字符集
 - ✓ 浏览器最终采用的字符集

- ✓ 解决乱码问题：

✓ 最优：

- ✓ 【编程】文本文件使用统一字符集，如UTF-8
- ✓ 【编程】正确填写meta。
- ✓ 【服务器设定】Content-Type不设定字符集

✓ 次优：

- ✓ 【编程】正确填写meta。
- ✓ 【服务器设定】Content-Type不设定字符集

动手实验 10-3

JSON

知识点 10-9

- ✓ JSON 指的是 JavaScript 对象表示法
(JavaScript Object Notation)
- ✓ JSON对象 串行化
持久化
 - ✓ {}中书写
 - ✓ 名称/值对, “key” : “value”
- ✓ JSON数组
 - ✓ []中书写
 - ✓ 对象间, 用逗号分隔

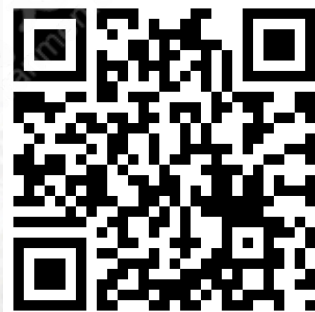
JSON

动手实验 10-4

- ✓ string -> JSON
- ✓ JSON -> string

欢迎大家线下与我沟通

✓ 微信群



扫码此二维码进入分享群
可获得更多**书单**和**学习资料**