

FIG.1

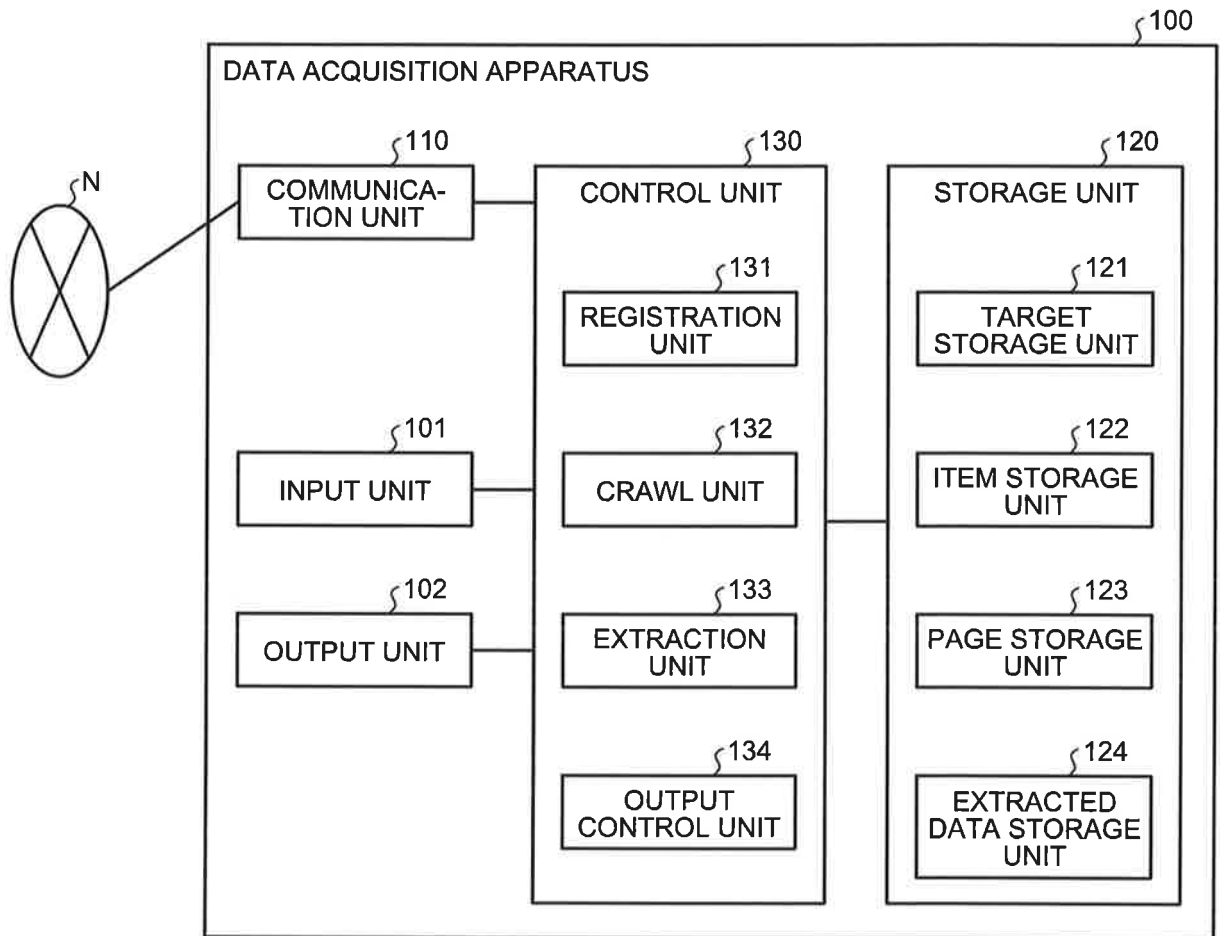


FIG.2

121

URLID	TARGET URL	POSITION IDENTIFICATION INFORMATION OF PART TO BE EXTRACTED		
		TITLE	ADDRESS	...
1	http://aaaa.bbb.ccc/ddd/eee/001.html	<DIV class="title"> </DIV> ORDERING SEQUENCE NUMBER: 1, /title/	<DIV class="address"> </DIV> ORDERING SEQUENCE NUMBER: 1, /info/address/	...
2	http://aaaa.bbb.ccc/ddd/eee/002.html	<DIV class="title"> </DIV> ORDERING SEQUENCE NUMBER: 2, /tps/table/	/<DIV.*>(<.+>)</DIV>/ /ADDRESS: (<.+>)\$/, ORDERING SEQUENCE NUMBER: 3, /info/address/	...
3	http://aaaa.bbb.ccc/ddd/eee/003.html	div#left h2, ORDERING SEQUENCE NUMBER: 3, /tps/table/	#infoContent @ LOCATION \s+? (<.+>) @is, ORDERING SEQUENCE NUMBER: 5, /info/address/	...
...	...	...	...	...

FIG.3

122


ITEM ID	DATA NAME	DATA TYPE	SEGMENTING TECHNIQUE	...
1	TITLE	CHARACTER	CSS SELECTOR	...
2	ADDRESS	CHARACTER	CSS SELECTOR, REGULAR EXPRESSION	...
3	TELEPHONE NUMBER	NUMBER	CSS SELECTOR, REGULAR EXPRESSION	...
4	DATE OF UPDATE	DATE	REGULAR EXPRESSION	...
5	POSITIONAL INFORMATION	LATITUDE AND LONGITUDE	CSS SELECTOR, REGULAR EXPRESSION	...
6	DESCRIPTION	CHARACTER	CSS SELECTOR	...
...	...	...	...	...

FIG.4

123

URLID	TARGET URL	STORAGE AREA	...
1	http://aaaaa.bbb.ccc/ddd/eee/001.html	D:\page\aaaaa.bbb.ccc\ddd\eee\001	...
2	http://aaaaa.bbb.ccc/ddd/eee/002.html	D:\page\aaaaa.bbb.ccc\ddd\eee\002	...
3	http://aaaaa.bbb.ccc/ddd/eee/003.html	D:\page\aaaaa.bbb.ccc\ddd\eee\003	...
...	...	...	...

FIG.5

124

URLID	TITLE	ADDRESS	TELEPHONE NUMBER	DATE OF UPDATE	POSITIONAL INFORMATION	DESCRIPTION	...
1	SANNAI-MARUYAMA SITE	xxx AZA MARUYAMA, SANNAI, AOMORI-SHI	017-781-xxxx	2014/10/20	N40° 81' 12" E140° 69' 56"	THE SANNAI-MARUYAMA SITE, A SPECIAL HISTORICAL SITE, IS ...	...
2	HAKKODA ROPEWAY	yyv AZA KANSUIZAWA, ARAKAWA, AOMORI-SHI	017-738-yyvy	2014/10/29	N40° 68' 08" E140° 83' 18"	THE HAKKODA ROPEWAY IS ...	...
3	TSUGARU-HAN NEPUTA VILLAGE	zzz KAMENOKO-MACHI, HIROSAKI-SHI, AOMORI	0172-39-zzzz	2014/6/3	N40° 61' 15" E140° 46' 92"	THE TSUGARU-HAN NEPUTA VILLAGE IS ...	...
...	...	...	...	...	...	...	...

FIG. 6

[illegible]

FIG.7

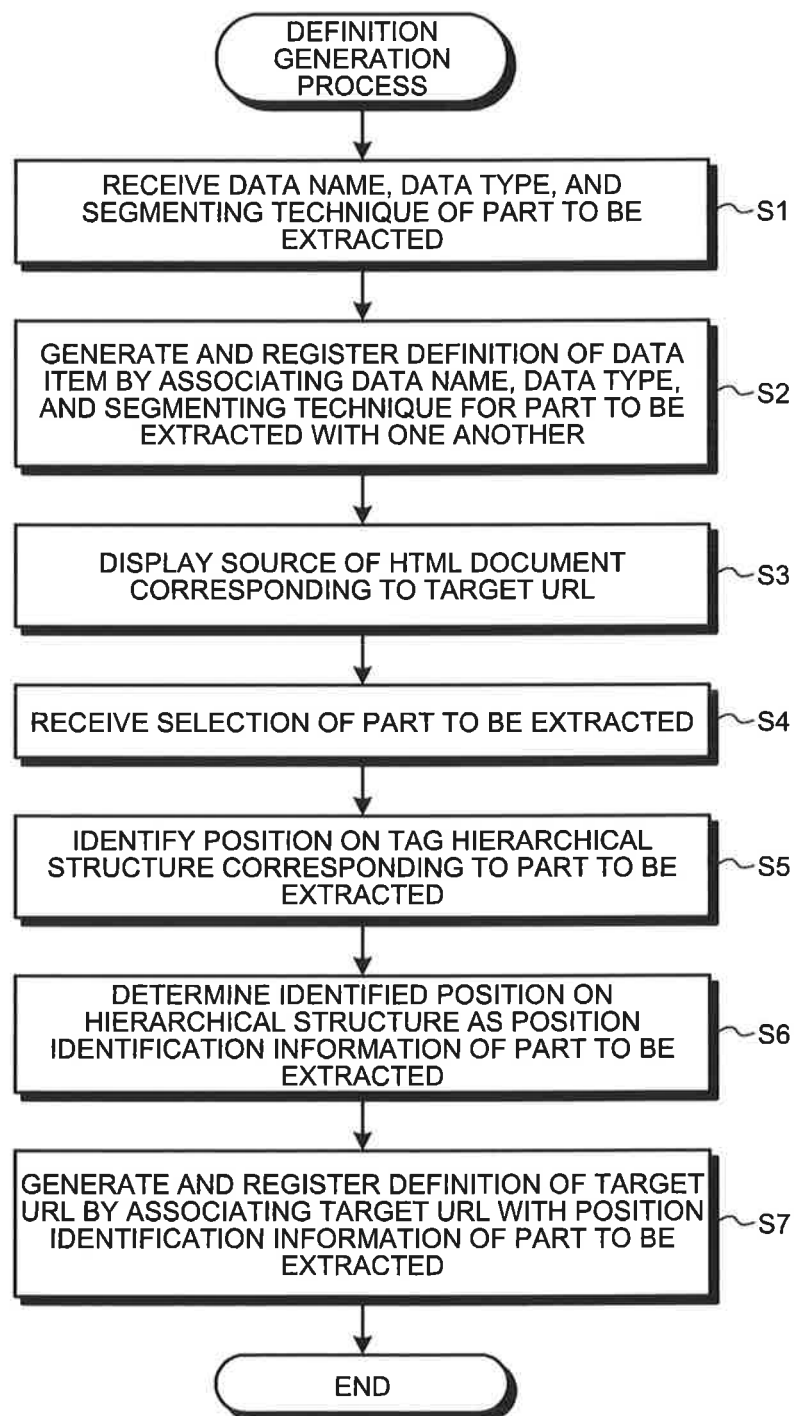


FIG.8

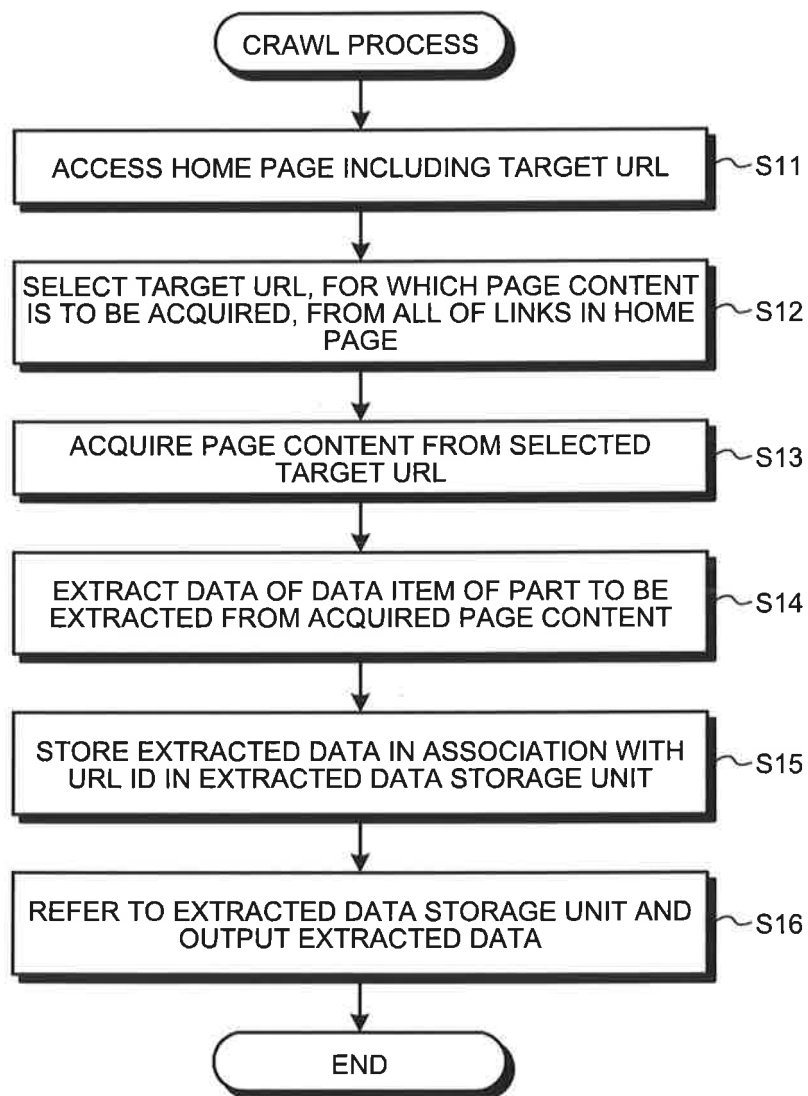


FIG.9

