

FIG.1

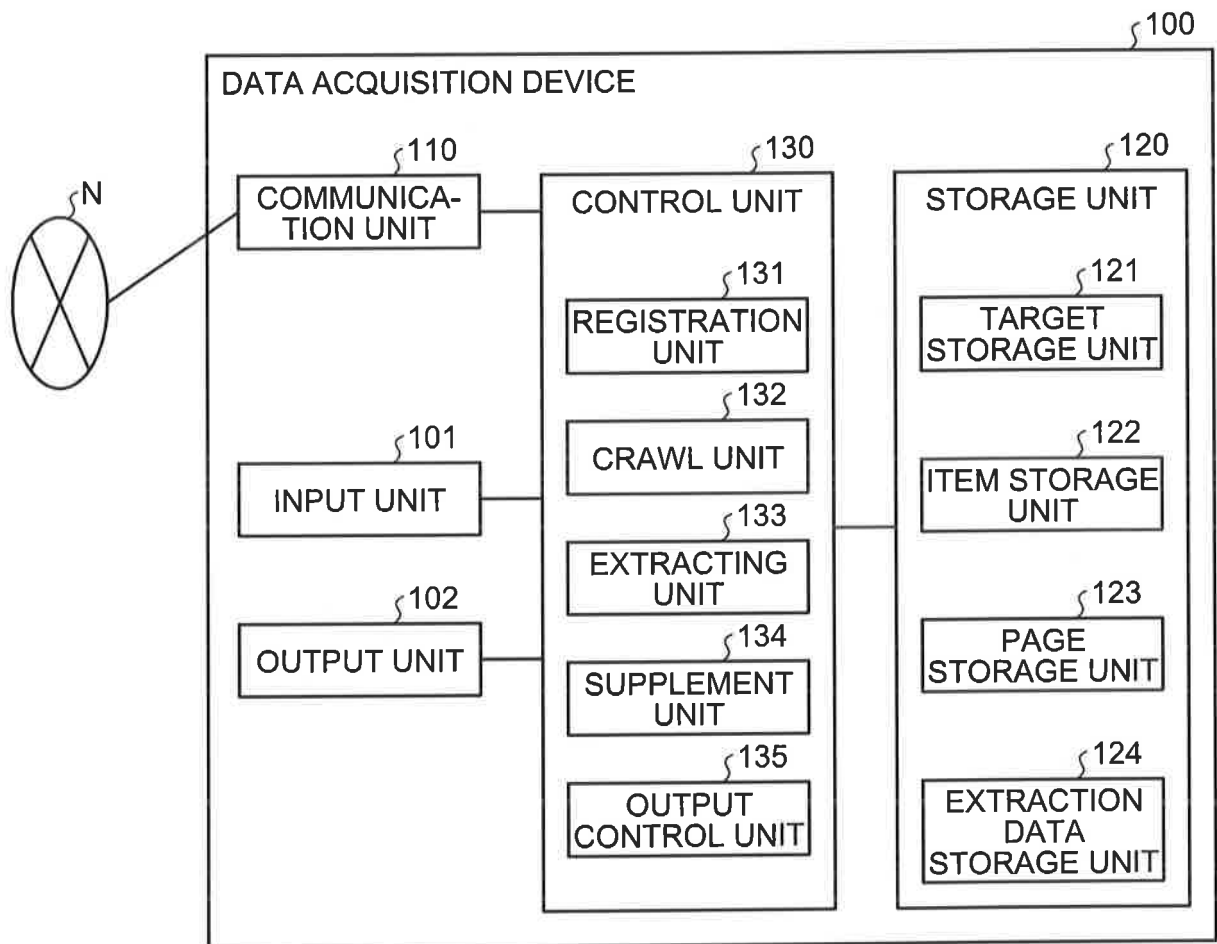


FIG.2

121



URLID	TARGET URL	POSITION SPECIFYING INFORMATION OF EXTRACTION TARGET PORTION			...
		TITLE	ADDRESS	...	
1	http://aaaa.bbb.ccc/ddd/eee/001.html	<DIV class="title"> </DIV> ORDER: 1, /title/	<DIV class="address"> </DIV> ORDER: 1, /info/address/	...	...
2	http://aaaa.bbb.ccc/ddd/eee/002.html	<DIV class="title"> </DIV> ORDER: 2, /tps/table/	/<DIV.*>(<.+>)</DIV>/ /ADDRESS:(<.+>)\$/ ORDER: 3, /info/address/	...	...
3	http://aaaa.bbb.ccc/ddd/eee/003.html	div#left h2, ORDER: 3, /tps/table/	#infoContent @<h3>LOCATION</h3>\$s+?<p>(<.+>?)</p>@is, ORDER: 5, /info/address/	...	...
...	...	...	...	...	...

FIG.3

122

ITEM ID	DATA NAME	DATA TYPE	SEGMENTATION METHOD	RELATION	...
1	TITLE	CHARACTER	CSS SELECTOR	ADDRESS, PHONE NUMBER	...
2	ADDRESS	CHARACTER	CSS SELECTOR, REGULAR EXPRESSION	FACILITY NAME, PHONE NUMBER, POSITIONAL INFORMATION	...
3	FACILITY NAME	CHARACTER	CSS SELECTOR, REGULAR EXPRESSION	TITLE, ADDRESS, PHONE NUMBER	...
4	PHONE NUMBER	NUMERAL	CSS SELECTOR, REGULAR EXPRESSION	TITLE, ADDRESS	...
5	UPDATE DATE	DATE	REGULAR EXPRESSION	TITLE, EXPLANATION	...
6	POSITIONAL INFORMATION	LATITUDE AND LONGITUDE	CSS SELECTOR, REGULAR EXPRESSION	ADDRESS, FACILITY NAME, PHONE NUMBER	...
7	EXPLANATION	CHARACTER	CSS SELECTOR	TITLE, UPDATE DATE	...
...	...	...	...	...	...

FIG.4

123

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

FIG.5

124

URLID	TITLE	ADDRESS	FACILITY NAME	PHONE NUMBER	UPDATE DATE	POSITIONAL INFORMATION	EXPLANATION	...
1	SPECIAL HISTORICAL SITE SANNAL-MARUYAMA SITE	AOMORI CITY, SANNAL-AZA MARUYAMA xxx	SANNAL-MARUYAMA SITE	017-781-xxxx	2014/10/20	N40°81'12" E140°69'56"	SPECIAL HISTORICAL SITE OF SANNAL-MARUYAMA SITE IS...	...
2	TOWADA-HACHIMANTAI NATIONAL PARK HAKKODA ROPEWAY	AOMORI CITY, ARAKAWA AZA KANSUIZAWA yy	HAKKODA ROPEWAY	017-738-yyyy	2014/10/29	N40°68'08" E140°83'18"	HAKKODA ROPEWAY IS...	...
3	TSUGARU-HAN NEPUTA-MURA HOMEPAGE	AOMORI-KEN, HIROSAKI CITY, KAMENOKO-MACHI zzz	TSUGARU-HAN NEPUTA-MURA	0172-39-zzzz	2014/6/3	N40°61'15" E140°46'92"	TSUGARU-HAN NEPUTA-MURA IS...	...
...	...	...	...	...	...	...	...	...

Figure 1 illustrates a data processing system (21) designed for extracting and processing data from a source. The system consists of a main processing area (23) and a source window (22).

The main processing area (23) is divided into several sections:

- SELECT EXTRACTION TARGET PORTION:** A dropdown menu for selecting the extraction target.
- PRIORITY:** A numeric input field for setting the priority of the extraction target.
- EXTRACTION DEFINITION:** A text input field for defining the extraction criteria.
- CONVERSION PROCESSING:** A text input field for specifying conversion rules.
- CHECK RESULT:** A status indicator showing the progress of the extraction and conversion process.

The source window (22) displays the raw data being processed, showing HTML code. A cursor (25) is positioned over a specific line of code, which is highlighted by a dashed box (27). The source code includes various HTML tags and attributes, such as <div>, <h3>, <p>, and .

FIG.7

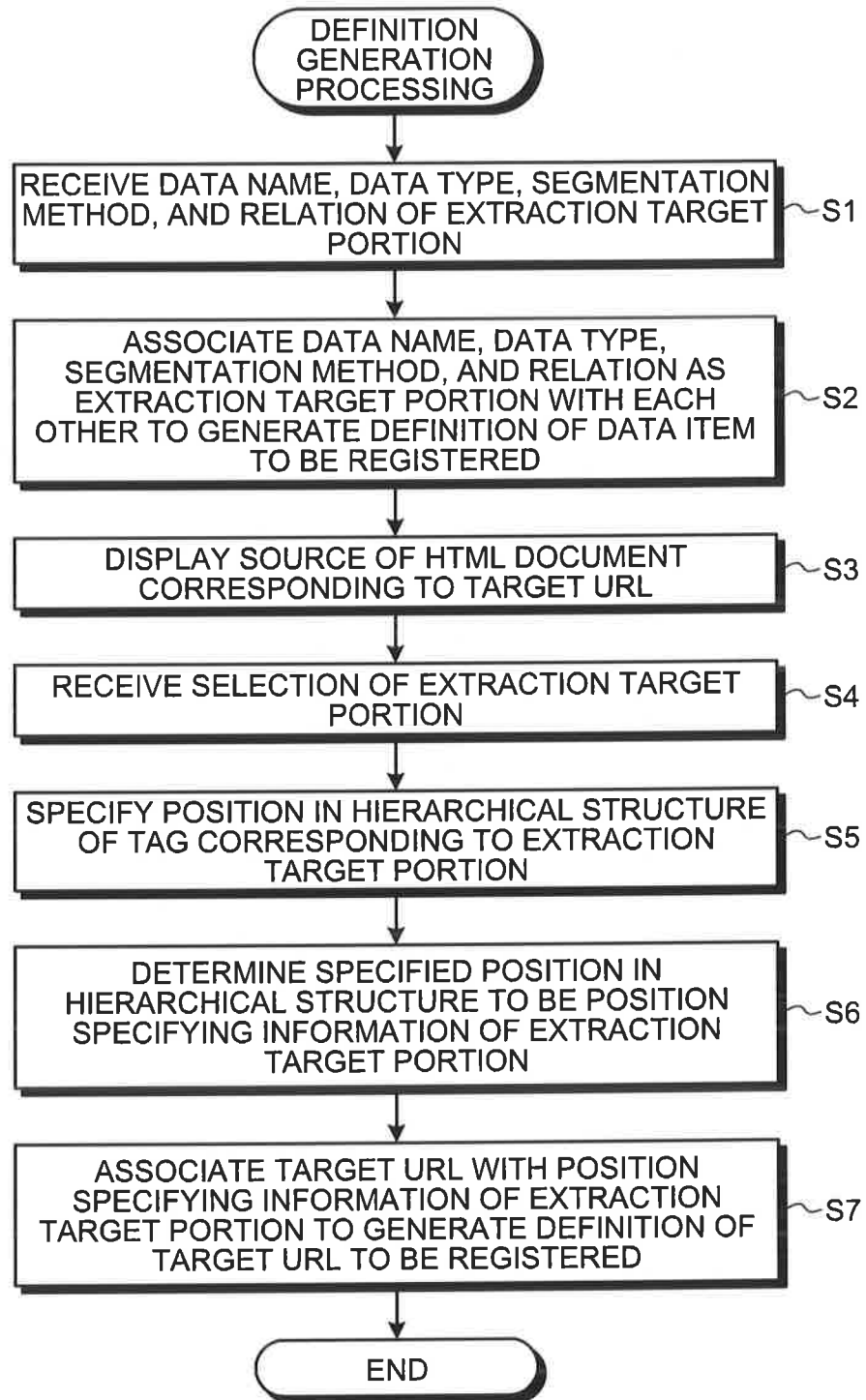


FIG.8

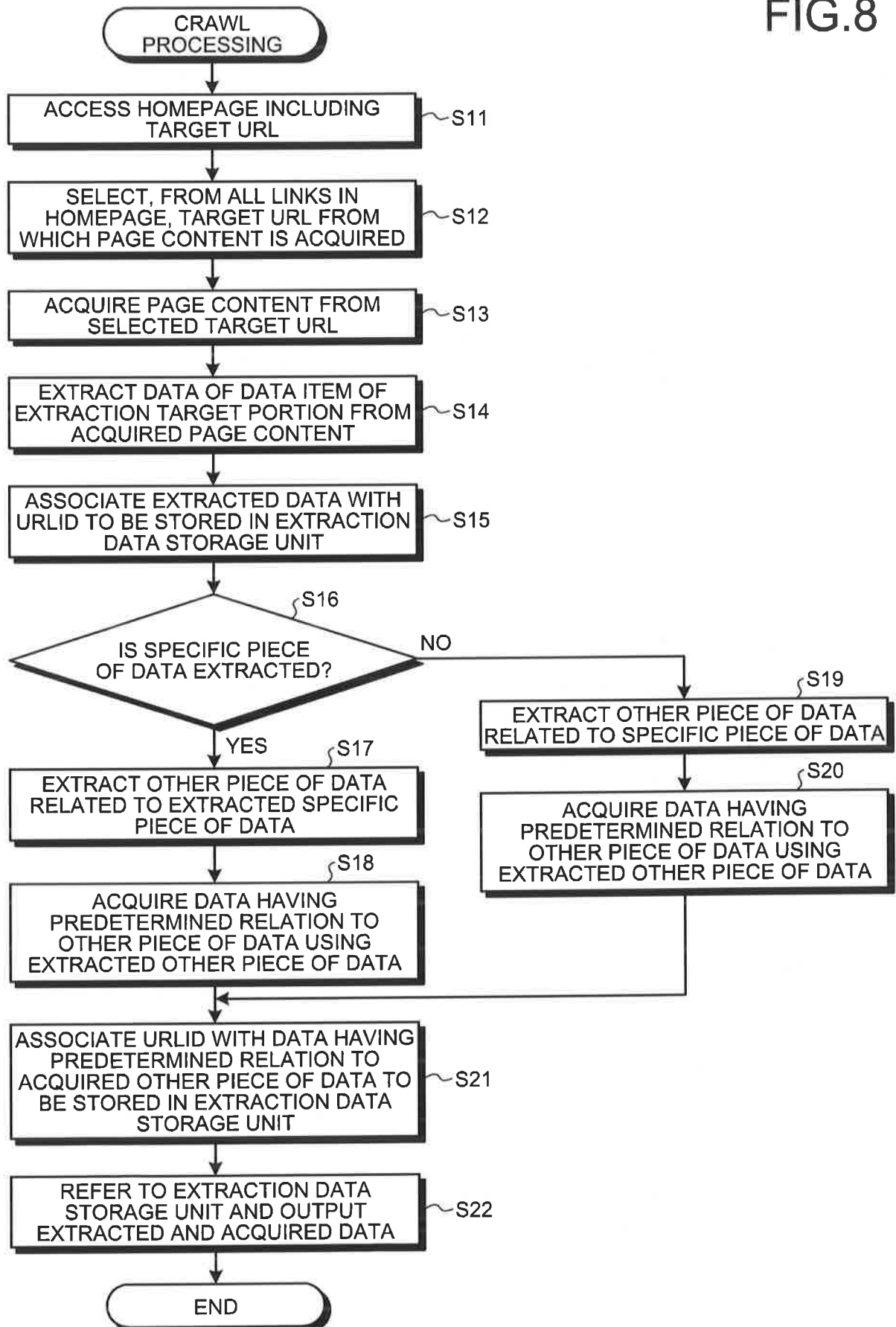


FIG.9

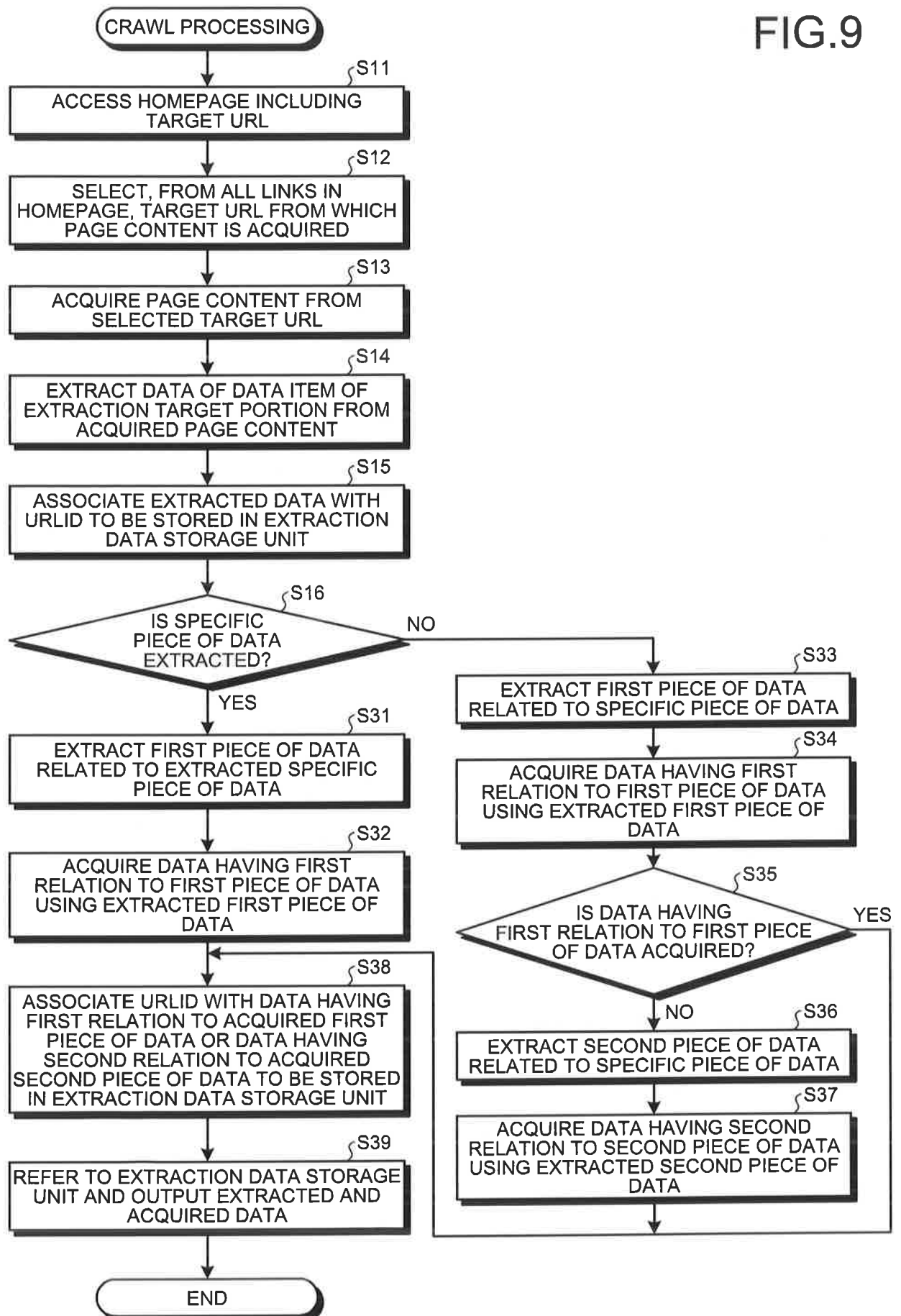


FIG.10

