Analysis of Factors Affecting the Filling rate of each type of daycare center Using Bigdata Technology

Analysis of Factors Affecting the Filling rate of each type of daycare center Using Bigdata Technology

Jeongwon Lee[1*], Byungil Jeon[1], Semin Kim[2], Gyujeon Lee[1], and Choong Ho Lee[1]

[1]Department of Information Communication Engineering, Hanbat National University, 34158, South Korea
[2]Department of Computer Education, Jeonju National University of Education, Jeonju 55101, South Korea
[Email : chlee@hanbat.ac.kr]

Abstract

Emphasizing the government's efforts to strengthen social responsibility and support for daycare center and daycare center for the growing number of working women, and to relieve the stress on the daycare center of working women as a result of the publicity of daycare center policies and the localization of daycare center have.

In this study, the current status of daycare centers and childcare services using public data centered on Seoul and Incheon is investigated and the factors of type of daycare center that affect the filling rate are derived using bigdata technology.

Index Terms: Daycare center, Daycare center fill rate, Data Analysis, Big Data Analysis

I. INTRODUCTION

In order to strengthen social responsibility and support for childcare and childcare for working women, this study was conducted to investigate the status of daycare and childcare services using public data to determine the appropriate daycare center based on scientific data, and to use Big Data technology to derive factors for each type of daycare center affecting the recruitment rate.

II. Analysis Procedure

In order to derive factors by type of day care center, we conducted analysis by data collection and storage, data processing, data analysis and visualization.

A. Data Collection and Storage

1) Data Collection

a) Data extractin technique
• Web Crawling

```
public class ChildCareDetailUrlProperties {
    // ----- 어린이집상세정보 기본 URL -----
    private static final String  DETAIL_URL      = "http://info.childcare.go.kr/info/pnis,

    // ----- JSP 페이지 정보 -----
    // [HTML]22_어린이집상세-기본현황
    private static final String  DETAIL_BASIS_JSP  = "BasisPresentConditionS1Pu.jsp";
    // [HTML]23_어린이집상세-영유아 및 교직원
    private static final String  DETAIL_STAFF_JSP  = "ChildStaffS1Pu.jsp";
    // [HTML]24_어린이집상세-교육보육과정
    private static final String  DETAIL_CARE_JSP   = "ChildCareCurriculumS1Pu.jsp";
    // [HTML]28_어린이집상세-평가인증
    private static final String  DETAIL_AUTH_JSP   = "AppraisaAuthenticationS1Pu.jsp";

    // ----- flag 정보 -----
    // [HTML]22_어린이집상세-기본현황
    public static final String   FLAG_DETAIL_BASIS  = "GH";
    // [HTML]23_어린이집상세-영유아 및 교직원
    public static final String   FLAG_DETAIL_STAFF  = "BG";
    // [HTML]24_어린이집상세-교육보육과정
    public static final String   FLAG_DETAIL_CARE   = "BB";
    // [HTML]28_어린이집상세-평가인증
    public static final String   FLAG_DETAIL_AUTH   = "PI";
```

b) Data collection and storage techniques
• Kafka : Managing messages for distributed processing
• Flume : Save Kafka Topinc messages to HDFS

```
# Agent name 설정
Agent_NurseryListInfo.sources = Source_kafka
Agent_NurseryListInfo.sinks = Sink_hdfs
Agent_NurseryListInfo.channels = Channel_file

# sources 설정
Agent_NurseryListInfo.sources.Source_kafka.type = org.apache.flume.source.kafka.KafkaSource
Agent_NurseryListInfo.sources.Source_kafka.zookeeperConnect = 192.168.10.4:2181,192.168.10.5:2181,192.168.10.6:2181
Agent_NurseryListInfo.sources.Source_kafka.topic = NurseryListInfo
Agent_NurseryListInfo.sources.Source_kafka.groupId = nurseryKafkaFlume
Agent_NurseryListInfo.sources.Source_kafka.batchSize = 100
Agent_NurseryListInfo.sources.Source_kafka.kafka.consumer.timeout.ms = 100

# sinks 설정 : hdfs
Agent_NurseryListInfo.sinks.Sink_hdfs.type = hdfs
#Agent_NurseryListInfo.sinks.Sink_hdfs.hdfs.path = hdfs://hdp01.Tech1.KoDB.kr/nursery/flume/hive/NurseryListInfo/%y-%m-%d
Agent_NurseryListInfo.sinks.Sink_hdfs.hdfs.path = hdfs://hdp01.Tech1.KoDB.kr/nursery/flume/hive/NurseryListInfo
Agent_NurseryListInfo.sinks.Sink_hdfs.hdfs.fileType = DataStream
Agent_NurseryListInfo.sinks.Sink_hdfs.hdfs.writeFormat = Text
Agent_NurseryListInfo.sinks.Sink_hdfs.hdfs.filePrefix = kafka_data
Agent_NurseryListInfo.sinks.Sink_hdfs.hdfs.fileSuffix = .dat
Agent_NurseryListInfo.sinks.Sink_hdfs.hdfs.threadsPoolSize = 10
Agent_NurseryListInfo.sinks.Sink_hdfs.hdfs.rollCount = 100
Agent_NurseryListInfo.sinks.Sink_hdfs.hdfs.rollInterval = 30
Agent_NurseryListInfo.sinks.Sink_hdfs.hdfs.roundValue = 5
Agent_NurseryListInfo.sinks.Sink_hdfs.hdfs.roundUnit = minute

# channel 설정 - file
Agent_NurseryListInfo.channels.Channel_file.type = file
Agent_NurseryListInfo.channels.Channel_file.checkpointDir = /usr/hdp/current/flume-server/channels/NurseryListInfo/checkpoint2
Agent_NurseryListInfo.channels.Channel_file.dataDirs = /usr/hdp/current/flume-server/channels/NurseryListInfo/data2
Agent_NurseryListInfo.channels.Channel_file.capacity = 100000000
Agent_NurseryListInfo.channels.Channel_file.transactionCapacity = 100000

# channel에 sources와 sinks 바인딩
Agent_NurseryListInfo.sources.Source_kafka.channels = Channel_file
Agent_NurseryListInfo.sinks.Sink_hdfs.channel = Channel_file
```

c) Data collection and Upload
• Download demographic data

B. Data Processing
1) Data processing methods and usage techniques
a) Spark SQL : Generate and summarize derivative data

b) Spark ML : association analysis (FPGrowth) utilization

c) association analysis (FPGrowth)
• Subordination variables : crperson_rate_gt_90_yn(whether the recruitment rate is over 90%)

C. Data Analysis and Visualization

1) Data Visualization Methods and Usage Techniques
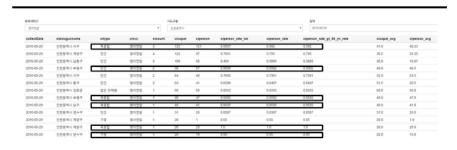#### a) data visualization technique : Zepplin

III. RESULTS

A. Incheon City Association analysis

- More than 90% of the national and infant-only type of daycare centers
- In the case of special day care centers for children with disabilities, the rate of higher recruitment is over 90%
- The western part of Incheon city has a high recruitment rate

B. Incheon city Association Analysis Result

- The Status of Childcare Centers in Incheon City

IV. CONCLUSION

There are 8 districts and 2 counties under Incheon city, and there are only 5 daycare centers that provide specialized services for children with disabilities

Among the daycare centers that provide infant care services, national daycare centers have a high recruitment rate of 97% ~ 98%

REFERENCES

[1] Kim Jae-hwan, a researcher on the public standards and evaluation of the Seoul-style daycare center, and the study on national education and childcare administration, 13 (4), 141-173, 2009.

[2] Ministry of Health and Welfare, Public Nursery Work Manual for each year, 2009.

[3] The Ministry of Health and Welfare. Press release: "The way that private day care centers become excellent day care centers! - Public and autonomous childcare project plan public hearings –" (Drained from http://www.mohw.go.kr), 2011.

[4] Seo Moon-hee, Yang Mi-sun, and Domi-hyang Transmission. Monitoring and Evaluation of Public-type and Autonomous Nursery School Pilot Projects. Ministry of Health and Welfare, Childcare Policy Research Institute, 2011.

[5] Seo Moon-hee, Yoo Hae-mi, Kim Jong-sook Shin Yoon-jung Yoon Jae-young Kim Hye-jin, 2012.