

공공 빅데이터를 사용한 자전거 교통 시스템 설계 및 활용

송병준*, 최현호^o, 손주희**, 문유진**

**한국외국어대학교 경영정보학과

^o한국외국어대학교 GBT학부

e-mail: sbjcap@gmail.com*, {777655256O, wngmlson**}@naver.com, yjmoon@hufs.ac.kr**

Building and Application of Bicycle Transportation System Utilizing Public Big Data

Byoung-Jun Song*, Hyun-Ho Choi^o, Ju-Hee Son**, Yoo-Jin Moon**

**Dept. of MIS, HanKuk University of Foreign Studies

^oDept. of GBT, HanKuk University of Foreign Studies

● 요약 ●

이 논문은 공공 빅 데이터를 활용하여 ‘서울자전거 따릉이’를 기반으로 자전거 교통 데이터베이스 시스템을 제작하고, 제작한 데이터베이스 시스템을 활용하는 예시를 보여주고 있습니다. 데이터베이스 시스템의 제작 과정을 통하여 데이터베이스 설계, 데이터 수집, 제작 및 가공 과정, 데이터베이스 시스템의 유용한 활용 예시를 확인할 수 있습니다. 버스, 택시와 같이 따릉이와 연계할 수 있는 다양한 대중교통 데이터를 추가로 활용한다면 더욱 정확하게 따릉이의 발전방향과 혼잡한 교통 환경 개선을 제시할 수 있는 유용한 정보를 도출할 수 있을 것이다.

키워드: 따릉이(Ttareungyi), 자전거(Bicycle), 데이터베이스(Database), 대중교통(Public Transportation), 공공 빅데이터(Public Big Data)

I. Introduction

서울시는 2015년 10월, 5대 권역(여의도, 상암, 신촌, 4대문 안, 상수)에 따릉이 대여소 150개소와 자전거 2000대 운영을 시작으로, 2018년 현재 약 2만 대까지 그 운영 규모를 확장하였으며, 시민들의 높은 호응에 힘입어, 2020년까지 운영대수를 4 만대까지 늘릴 계획이다. 하지만 따릉이의 운영 규모가 확대될수록 따릉이에 대한 불만 사항 또한 많아지기 시작하였다. 이 과정에서 따릉이와 관련된 누적 데이터를 확보하고 가공하여 사실에 기반한 유용한 정보를 도출하여 활용할 수 있도록, 이 논문은 더욱 효과적인 개선 방안을 제시할 것이다. 이 논문에서는 기존 연구에서 도출한 연구 결과와 따릉이와 관련된 실제 빅데이터 즉 대여 이력과 자전거 도로 등을 기반으로 데이터베이스를 구축하고 활용하는 과정을 담고 있다.

자전거 재배치 관리 구역을 설정하고 구역 내 자전거 재배치를 위한 운송 수단의 경로를 해결하는 문제를 다루었다.[2] 상기 연구들은 공공자전거 관련 데이터를 연계하고 활용하는 방안을 제시하였으며, 데이터간의 연계를 통하여 교통 시스템에 더욱 유용한 정보를 산출할 수 있다는 것을 보여준다.

2. Related Database Design

장영관 외 (1996) 연구에서는 관계형 데이터베이스를 비정규화를 사용하여 데이터베이스를 설계하는 방법을 제시하였다.[6] 해당 연구는 저장용량에 제한이 있는 데이터를 정규화 과정을 일부 생략하여도 유연성 있는 데이터베이스를 생산할 수 있도록 하였다.

II. Preliminaries

1. Related Works

Vogel 외 (2011) 연구에서는 오스트리아의 자전거 대여 서비스를 이용자의 사용패턴으로 시각화 하고 지리학적 정보로 표현하였다.[1] Schuijbroek 외 연구 (2017)에서는 사용자 서비스 수준을 고려하여

III. System Architecture

1. E-R Model

제작된 데이터베이스의 시스템 구성은 다음과 같다. 테이블은 Borough(서울시 구 목록), RentalCenter(대여소 정보), RentalLog

(대여이력), TRI(자전거 번호) RentalType(대여구분), Citizen_Satisfaction(시민 만족도), BikeRoad(자전거도로) 총 7개로 구성되어 있으며, 추가 제작된 지하철 거리 관련 데이터는 RentalCenter 테이블에 삽입되었다.

2. Processing of the Database System



Fig. 1. Java를 이용한 CSV 데이터 XLS 분할 과정

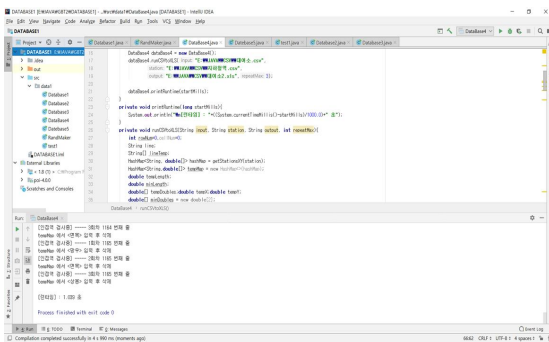


Fig. 2. Java를 이용한 대여소, 지하철 직선거리 좌표계산

그림 1, 2는 지하철 거리 관련 데이터를 가공하고 기존 대여소 데이터에 삽입하는 과정을 보여준다.

3. Representative Useful Information

3.1 각 대여소별 지하철역 거리

그림 3은 각 대여소와 인접한 지하철역 목록, 지하철역 사이 거리를 산출한 결과이다. 해당 유용한 정보는 시민들의 자전거 사용을 구하는 데 활용될 수 있을 것이다.

대여소번호	대여소명	인접지하철역1	인접거리1(M)	인접지하철역2	인접거리2(M)	인접지하철역3	인접거리3(M)	인접지하철역4	인접거리4(M)	인접지하철역5	인접거리5(M)
1	101	(구)합정동 주민센터	합정	합정	798	마포구청	1416				
2	102	합정역 1번출구 앞	합정	합정	41	마포구청	1034				
3	103	합정역 2번출구 앞	합정	합정	138	마포구청	1137				
4	104	합정역 1번출구 앞	합정	합정	170	합정	736	상수	836		
5	105	합정역 2번출구 앞	합정	합정	121	합정	773	상수	884		
6	106	합정역 1번출구 앞	합정	합정	722	합정	794	상수	1008		

Fig. 3 각 대여소별 지하철역 거리

3.2 출근시간대 자전거 이용률

```
SELECT COUNT(CASE WHEN DATEPART(HH,대여일시)
>='06' AND DATEPART(HH,대여일시)<'09' THEN 1
END)*100.0 / COUNT(대여일시) AS [출근시간대비율(%)]
COUNT(CASE WHEN DATEPART(HH,대여일시)>='17' AND
DATEPART(HH,대여일시)<'20' THEN 1 END)*100.0 /
COUNT(대여일시) AS [퇴근시간대비율(%)]
COUNT(CASE WHEN (DATEPART(HH,대여일시)>='06'
AND DATEPART(HH,대여일시)<'09') OR (DATEPART(HH,
대여일시)>='17' AND DATEPART(HH,대여일시)<'20')
THEN 1 END)*100.0 / COUNT(대여일시) AS [출퇴근시간대
비율(%)]
FROM 대여이력;
```

IV. Conclusions

본 논문에서는 서울시 공공자전거와 관련된 데이터를 가공하고 데이터베이스를 구축하여 활용하는 방법을 제시하였다. 이를 통하여 따릉이, 서울시 자전거 이용환경 및 서울시 교통 환경 개선에 이용할 수 있는 유용한 정보를 도출할 수 있다는 결과를 보여주고 있다. 차후 논문에 제시된 데이터 예시뿐 아니라 버스, 택시와 같이 따릉이와 연계할 수 있는 다양한 대중교통 데이터를 추가로 활용한다면 더욱 정확하게 따릉이의 발전방향을 제시할 수 있는 유용한 정보를 도출할 수 있을 것이다.

REFERENCES

[1] Vogel, P., T. Greiser, and D.C. Mattfeld, "Understanding bike-sharing systems using Data Mining : Exploring activity patterns," Procedia Social and Behavioral Sciences, vol.20, pp514-523, 2011.

[2] Schuijbroek, J., R.C. Hampshire, and W.-J. van Hoeve, "Inventory rebalancing and vehicle routing in bike sharing systems," European Journal of Operational Research, vol.257, no.3, 2017.