

Data Fragmentation Protection Technique for the Performance Enhancement of DB-Based Navigation Supporting Incremental Map Update

Yong Ho Kim[†] · Jae Kwang Kim^{††} · Seongil Jin^{†††}

ABSTRACT

Most of the navigation in the vehicle has been developed based on a complex structure of PSF(Physical Storage Format) files, making it difficult to support incremental map updates. DB-based navigation is drawing attention as a next-generation navigation method to solve this problem. In DB-based navigation that supports incremental map updates, data fragmentation due to continuous map data updates can increase data access costs, which can lead to a decrease in search performance. In this paper, as one of the performance enhancement methods of DB-based navigation that supports incremental map updates, data fragmentation prevention techniques were presented and the performance improvement effect was verified through actual implementation.

Keywords : Incremental Map Update, DB-based Navigation, Data Fragmentation

점증적인 맵 갱신을 지원하는 DB 기반 내비게이션의 성능 향상을 위한 데이터 단편화 방지 기법

김 용 호[†] · 김 재 광^{††} · 진 성 일^{†††}

요 약

차량에 탑재된 내비게이션의 대부분은 복잡한 구조의 PSF(Physical Storage Format) 파일 기반으로 개발되어 점증적 맵 갱신을 지원하기 어렵다. 이를 해결하기 위한 차세대 내비게이션 방법의 하나로서 DB 기반의 내비게이션 기술이 주목받고 있다. 점진적 맵 갱신을 지원하는 DB 기반 내비게이션 구현에 있어 지속적인 맵 데이터 갱신으로 인한 데이터 단편화현상으로 데이터 접근 비용이 증가할 수 있어 검색 성능의 저하가 발생할 수 있다. 본 논문에서는 점증적 맵 갱신을 지원하는 DB 기반 내비게이션의 성능 향상 방법의 하나로 데이터 단편화 방지 기법을 제시하고 실제 구현을 통하여 성능 향상 효과가 있음을 검증하였다.

키워드 : 점증적 맵 갱신, DB 기반 내비게이션, 데이터 단편화

1. 서 론

현재 대부분의 차량 내장형 내비게이션은 on-board 내비게이션으로 모든 데이터를 PSF(Physical Storage Format) 파일 형태로 저장하는 방식으로서 최소한의 용량과 우수한 검색 성능을 제공한다. 그러나 맵의 갱신 시 예를 들어 길 데이터가 하나만 바뀌더라도 전체 데이터를 바꿔야 하는 full update 방식으로 인터넷 또는 오프라인으로 전체 맵을 갱신하므로 비용 측면에서도 비효율적이며 무선망을 이용한 실시간 맵 갱신에 어려움이 있는 단점이 존재한다.

이를 해결하기 위한 점진적 맵 갱신 기술은 차세대 내비게이션 기술의 하나로서 인포테인먼트 서비스의 중요한 요소이다.

최근 점진적 맵 갱신을 지원하기 위해 파일 기반이 아닌 DB 기반 내비게이션 기술이 주목 받고 있다. 그동안 유럽 컨소시엄의 ACTMAP 프로젝트, NDS 컨소시엄의 시범프로젝트를 통해 상용화 연구가 진행되어 왔으며 국내에서는 (주)현대엠앤소프트와 (주)리얼타임테크가 협력하여 연구개발 및 상용화를 진행하여 왔다[1,3,4].

DB 기반 내비게이션은 점증적 맵 갱신을 유연하게 할 수 있는 기술이지만 상용화 관점에서 파일 기반 내비게이션의 용량과 성능에 큰 차이가 나지 않게 하는 여러 기술들을 해결하여야 한다. 또한 점진적 맵 갱신을 통해 차분 데이터를 DB에 갱신 시 데이터가 분산되는 데이터 단편화 현상을 해결하여야 한다.

[†] 준 회 원 : 충남대학교 컴퓨터공학과 석사과정

^{††} 비 회 원 : (주)리얼타임테크 내장형DBMS엔진팀 팀장

^{†††} 종신회원 : 충남대학교 컴퓨터공학과 교수/(주)리얼타임테크 대표

Manuscript Received : December 31, 2019

Accepted : January 23, 2020

* Corresponding Author : Seongil Jin(sijin@cnu.ac.kr)

점증적 맵 갱신을 위한 DB 기반 내비게이션에서 검색 연산 시 주변 구역에 있는 데이터를 MESH-ID 단위로 자주 검색하게 된다. 데이터 생성 초기 시에는 MESH-ID 단위로 페이지가 할당되나 추후 지속적인 점진적 맵 갱신 시 여러 페이지에 분산 적재되는 데이터 단편화 현상이 일어나 검색 연산 시 페이지 접근비용이 증가하여 초기 성능을 지속적으로 유지하지 못하게 된다. 본 연구에서는 이러한 데이터 단편화 현상을 해결하기 위해 맵 갱신 시 동일한 MESH-ID를 갖는 데이터가 동일한 페이지에 적재되어 클러스터링 되게 함으로써 접근비용을 최소화할 수 있도록 하는 MESH-ID 기반의 데이터 단편화 방지 기법을 제시하였다.

이러한 데이터 단편화 저장 기법을 ㈜리얼타임테크가 개발한 Kairos ite 모바일 DBMS[2]에 구현 적용하여 전국 단위 상용 데이터를 대상으로 시험한 결과 데이터 단편화에 따른 성능 저하를 해결하여 초기 성능이 지속적으로 유지될 수 있음을 검증하였다.

본 논문의 구성은 다음과 같다. 제2장에서는 파일 기반 내비게이션과 DB 기반 내비게이션에 대해서 기술하고, 제3장에서는 데이터 단편화 현상 및 문제점에 대해서 설명한다. 제4장에서는 본 논문에서 제시하는 MESH-ID 기반의 데이터 단편화 방지 기법에 대해 기술하고, 제5장에서는 시험 방법 및 환경에 대해 기술한다. 제6장에서는 시험에 대한 성능 평가를 하고, 제7장에서는 결론을 다룬다.

2. 파일 기반 내비게이션 VS DB 기반 내비게이션

2.1 파일 기반 내비게이션

거의 대부분의 기존 파일 기반 내비게이션에서 사용되고 있는 PSF 파일은 길, 배경, 디스플레이 등의 여러 데이터가 하나로 축약된 특별 파일 형태로서 용량이 작고 연산이 빠른 장점이 있다. 그러나 일부 데이터의 갱신 시 전체를 다시 빌드 할 수밖에 없어 무선망에서 실시간 점증적 맵 갱신 서비스를 제공할 수 없는 한계를 갖고 있어 차세대 내비게이션 서비스에 적합하지 못한 구조이다. 이러한 구조적 한계로 차량용 맵 갱신의 경우 수개월에 한 번씩 전체 맵 데이터를 다운로드 방식으로 변경할 수밖에 없으며 응용과 PSF 간의 결합도가 매우 강하므로 데이터에 대한 독립성과 다른 응용프로그램으로 확장성이 좋지 않은 원인이 되고 있다[5].

2.2 DB 기반 내비게이션

PSF 파일 기반 내비게이션의 단점인 점증적 맵 갱신을 지원하기 위해 하나의 축약된 파일 형태가 아닌 노드, 링크, 디스플레이, POI 검색용 데이터들을 테이블 형태로 분리 관리하는 DB 기반 내비게이션 기술이 차세대 기술로 주목받고 있다.

내비게이션 상용데이터는 업데이트 연산에 유연하게 작용할 수 있도록 데이터 유형별로 수개 또는 수십 개의 구성될 수 있다.

이러한 DB 기반 내비게이션은 데이터베이스의 테이블 구조를 사용하여 내비게이션 상용데이터를 관리하므로 파일 기반의 내비게이션에서는 불가능한 차분데이터만의 부분 업데이트 서비스 및 주행 중 업데이트를 제공할 수 있다.

또한 새로운 응용 프로그램의 개발과 확장이 용이하다. 그러나 DB 기반 내비게이션의 상용화를 위해서는 PSF 파일이 갖는 장점인 용량 문제와 검색 성능 문제 등을 해결해야 한다. 예를 들어 PSF 기반 내비게이션에는 색인구조가 따로 없고 물리적으로 구성된 레코드를 가져오면 되나 DB 기반 내비게이션에서는 DBMS의 복잡하고 다양한 색인 방법을 제공하여야 하며 성능 또한 사용자 관점에서 만족할 수 있는 수준이어야 한다[6,7].

3. 데이터 단편화 현상

내비게이션 상용데이터는 지도상의 구역을 MESH 단위로 구분한다. 구분된 구역은 MESH-ID로 쉽게 접근 가능하며 내비게이션 서비스 중 MESH-ID로 빈번하게 검색이 일어난다. 이를 위해 동일한 MESH-ID를 갖는 데이터는 클러스터링 되게 적재되어야 레코드 페이지에 대한 액세스 비용을 최소화할 수 있다.

기존 내비게이션용 모바일 DBMS[2]에서는 (Fig. 1)과 같이 입력되는 데이터의 MESH-ID를 기준으로 데이터 페이지에 순차적으로 적재되며 pointOfInsert가 가리키는 마지막 페이지에서 삽입이 되는 구조이다. 후에 점증적 맵 갱신이 일어나게 되면 데이터 저장관리의 효율성을 위해 우선적으로 레코드 페이지에 레코드의 삭제로 인한 빈 공간이 있는지를 검사하여 빈 공간이 있을 경우 pointOfInsert 페이지보다 우선적으로 삽입하게 되며 빈공간이 없을 경우에 pointOfInsert 페이지에 순차적으로 삽입하게 되어 동일한 페이지에 서로 다른 MESH-ID를 갖는 데이터가 적재되는 단편화 현상이 발생된다. 지속적으로 오랜 기간 점증적 맵 갱신이 일어나게 되면 이러한 현상은 더욱 심해져 데이터 페이지의 접근비용이 크게 증가하게 되어 심각한 성능 저하가 일어나게 된다[9].

이러한 점증적인 맵 갱신으로 인한 데이터 단편화 현상을 예를 들면 다음과 같다.

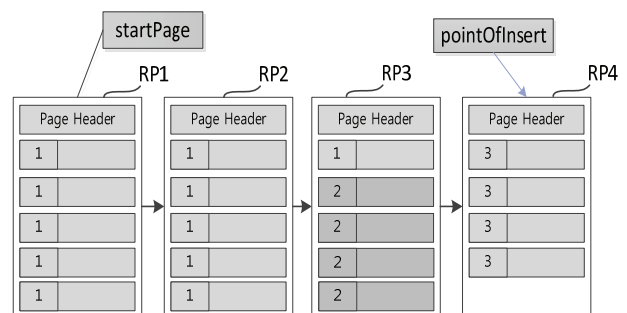


Fig. 1. Example of Record Page Structure

Fig. 1은 기존 DBMS의 레코드 페이지 저장 구조를 예시한 그림으로 RP1~RP4라는 4개의 데이터 페이지에는 각각 페이지 헤더가 상단에 있고, 나머지는 레코드 데이터가 적재된다. 각 레코드 데이터는 레코드의 앞부분에 MESH-ID의 값과 (Fig. 1에서는 편의상 숫자 1~3으로 표시) 데이터가 적재되는 구조이다. 최초 DB 생성 시 데이터의 적재는 레코드 페이지에 빈 공간이 존재하지 않는 형태이다.

이러한 구조하에서 맵 갱신 연산이 지속적으로 발생하게 되면 데이터는 Fig. 2와 같은 상황이 될 수 있다.

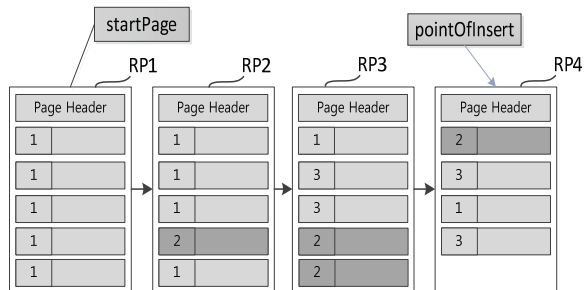


Fig. 2. Example of Fragmentation

짙은 색으로 표시되어 있는 MESH-ID 값이 2인 레코드는 최초에 RP3 레코드 페이지에 적재되어 있었는데, 지속적인 맵 갱신 연산으로 RP2에서 선행 삭제 연산으로 빈 공간이 생겨 MESH-ID 값이 2인 데이터가 적재되고 유사하게 RP3 대신에 RP4에 MESH-ID 값이 2인 데이터가 RP3에 삽입되어 단편화가 된다. 이로 인하여 MESH-ID 값이 2인 레코드 검색에 대해서 이전에는 한 개의 레코드 페이지를 접근하였는데, 맵 갱신 후에는 3개의 페이지에 대한 접근 비용이 발생하게 된다. 극단적인 경우 맵 갱신으로 인하여 MESH-ID 값이 2인 레코드가 4개의 레코드 페이지에 분산 저장된다.

4. MESH-ID 기반 데이터 단편화 방지 기법

본장에서는 데이터 단편화 현상을 발생시키는 구조적인 문제를 해결하기 위하여 해싱을 이용한 MESH-ID 기반 데이터 단편화 방지 기법을 제시한다.

4.1 MESH-ID 기반 해싱 저장구조

단편화 현상을 일으키는 기존 구조의 문제점은 점층적 맵 갱신에 따라 동일한 데이터 페이지에 서로 다른 MESH-ID 값을 갖는 레코드가 존재하는데 있다. 이를 방지하기 위해서 하나의 데이터 페이지에 동일한 MESH-ID 값만을 갖도록 MESH-ID 값을 해싱하여 동일한 해싱 값만을 갖는 데이터들에 대해서 데이터 페이지를 할당하여 관리하도록 하는 구조를 제안한다.

Fig. 3과 같이 데이터가 삽입될 때 MESH-ID 값을 해싱하여 관리하는 해싱 테이블을 유지하고 해싱 값과 연결된

(MESH-ID, startPage, pointOfInsert)값들을 갖는 메타 정보 레코드를 유지하며 충돌이 있을 시에는 이들 레코드를 연결 리스트 형태로 유지 관리한다. 이러한 레코드의 정보를 이용하여 데이터 페이지를 생성 관리하게 된다. 데이터 페이지 구조는 MESH-ID 값에 따라 서로 다른 데이터 페이지를 구성하는 것 이외에는 기존의 구조와 동일하다.

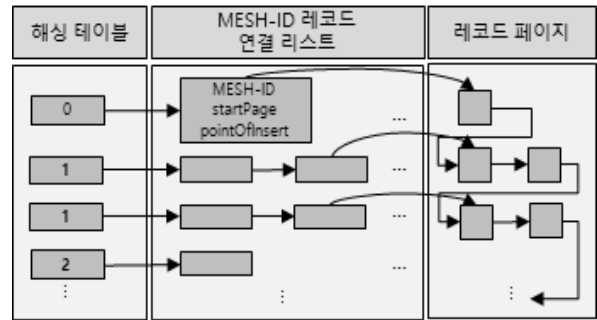


Fig. 3. Data Page Management Structure

제안한 구조는 해시 테이블과 연결 리스트 구조를 기존 구조에 추가하여 관리함으로써 레코드 적재 시 입력된 레코드의 MESH-ID를 기준으로 적재될 레코드 페이지를 구분하여 서로 다른 MESH-ID의 레코드가 동일한 페이지에 적재되지 않도록 한다[10-12].

이러한 구조를 유지하기 위한 추가적인 비용은 단편화 현상에 따른 데이터 접근비용 증가에 비하면 무시할 정도의 매우 미미한 수준이다.

4.2 단편화 방지 알고리즘

4.1에서 기술한 구조에 기반한 단편화를 방지하기 위한 알고리즘은 다음과 같다.

- 1단계 : 입력된 레코드 데이터의 MESH-ID 해싱 값을 구한다.
- 2단계 : 해시 테이블에 1단계 과정에서 구한 값이 존재하는지 조사하여 값이 없으면 새로운 MESH-ID 레코드를 생성하고 해싱 테이블과 연결한다. 이미 값이 존재하면 MESH-ID 레코드의 연결 리스트를 검색하여 삽입되는 MESH-ID 값을 갖는 MESH-ID 레코드가 없으면 레코드를 생성하여 기존 연결 리스트에 삽입하고 해싱 테이블과 연결한다.
- 3단계 : MESH-ID 레코드에 있는 삽입될 레코드 페이지에 관한 메타정보를 이용하여 레코드를 삽입한다.

상기와 같은 방법은 내비게이션 데이터의 점층적 맵 갱신 시마다 실행되며 하나의 페이지에 동일한 MESH-ID 값만을 갖도록 유지 관리할 수 있도록 한다.

5. 시험 방법 및 환경 소개

5.1 시험 방법

PSF 파일 기반 내비게이션을 DB 기반 내비게이션으로 전환하는 프로젝트를 진행하면서 개발한 ㈜리얼타임테크의 모바일 DBMS인 Kairos Lite에 점증적인 맵 갱신을 지원하는 DB 기반 내비게이션의 성능 향상을 위한 데이터 단편화 방지 기법을 구현 적용하여 다음과 같은 방법으로 시험을 진행하였다[8].

- 전국 단위의 상용 데이터 중 디스플레이 데이터인 POLYGON 테이블을 대상으로 시험을 수행.
- 사용한 POLYGON 테이블 스키마는 다음과 같으며 레코드의 수는 4,040,000 개로 전국 단위 상용 데이터임.

Field name	Type	Index	Key
ID	UNSIGNED INT	B+TREE	PRIMARYKEY
MESHID	UNSIGNED INT	B+TREE	
GEOMETRY	POLYGON	MLGRID	

- 기존 내비게이션용 모바일 DBMS의 문제점인 데이터 단편화 현상을 나타내기 위하여 시험 대상 테이블인 POLYGON 테이블의 데이터를 임의로 단편화시킴.
- 데이터 단편화 현상 정도를 표현하기 위해 클러스터율이라는 개념을 도입. (특정 MESH-ID 값을 가지는 레코드 데이터가 100개가 있을 경우 그중 10개의 레코드 데이터가 각각 서로 다른 레코드 페이지에 적재되었다면 클러스터율은 10%라 정의함)
- 클러스터율이 0인 경우는 데이터 단편화 현상이 발생하지 않은 것을 의미함. 클러스터율이 10%인 경우는 클러스터율이 0%인 DB보다 해당 MESH-ID의 레코드 수를 기준으로 10%의 페이지 접근 비용이 더 발생함을 의미함.
- 클러스터율을 0%~40%로 대상 테이블인 POLYGON 테이블의 데이터를 단편화하여 데이터를 적재함.
- 시험은 임의로 생성된 0%~40%의 대상 DB의 POLYGON 테이블에 특정 MESH-ID로 검색을 하여 접근되는 레코드 페이지 액세스 비용(Read Page Count)과 검색 시간(Search speed)을 측정함.
- 클러스터율이 0%~40%의 5세트의 DB의 POLYGON 테이블에 다음 6개 MESH-ID 샘플로 검색 연산을 실행함. (1955629063, 1955872775, 1955874949, 1956219974, 1956415682, 1964836967)
- MESH-ID 기반 데이터 단편화 방지 기법이 적용된 DBMS와 적용되지 않은 기존의 DBMS와의 검색 연산 시 레코드 페이지 액세스 비용(Read Page Count)과 검색 시간(Search speed)을 비교함.

5.2 시험 환경

5.1에서 서술한 시험 수행을 위한 시험 환경은 다음과 같다.

- CPU : Intel(R) Core(TM) i5-2500 CPU @3.30GHz
3.30GHz
- MEMORY : 8.0GB
- 운영체제 : Windows 7, 64 비트
- DBMS : Kairos Lite

6. 성능 평가

앞의 5장에서 서술한 시험 방법으로 기존 DBMS와 본 연구에서 제시한 데이터 단편화 방지 기법이 구현된 DBMS로 클러스터율이 0~40%인 각 5개의 POLYGON 테이블에 대해 특정 MESH-ID로 검색했을 시, 레코드 페이지 액세스 비용(Read Page Count)과 검색 시간(Search speed)을 측정 한 결과는 다음과 같다.

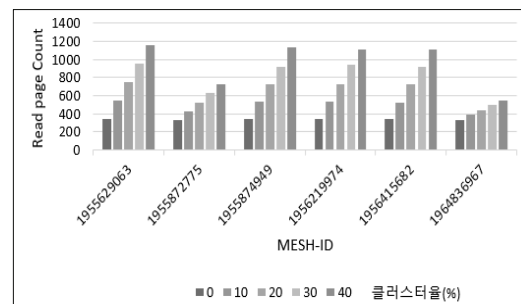


Fig. 4. MESH-ID Read Page Count According to Clustering Rate of Existing DBMS

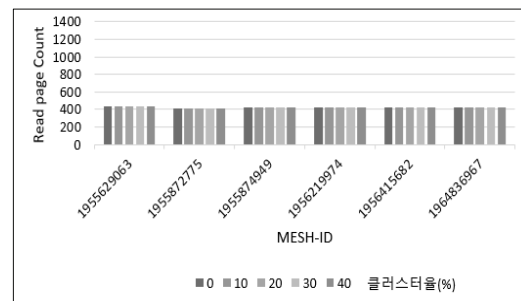


Fig. 5. MESH-ID Read Page Count According to Clustering Rate of DBMS with Clustering Technique

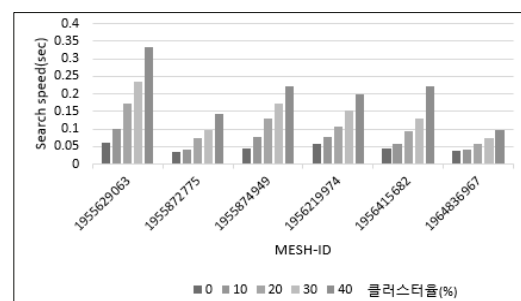


Fig. 6. MESH-ID Search Speed According to Clustering Rate of Existing DBMS

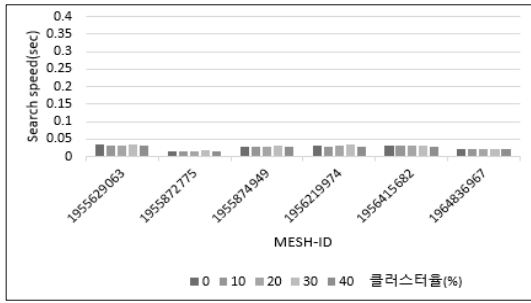


Fig. 7. MESH-ID Search Speed According to Clustering Rate of DBMS with Clustering Technique

Fig. 4, Fig. 6을 보면 기존의 DBMS는 클러스터율이 0%~40%로 증가함에 따라 검색 속도(Search speed) 및 레코드 페이지 액세스 비용(Read page count)이 증가함을 확인할 수 있다. 즉 데이터 단편화 현상이 발생하여 단편화 정도에 따라 데이터 페이지 접근 비용이 증가함을 볼 수 있다.

하지만 MESH-ID 기반 데이터 단편화 방지 기법이 구현된 DBMS의 시험 결과인 Fig. 5, Fig. 7을 보면 클러스터율이 변화하여도 검색 속도(Search speed) 및 레코드 페이지 액세스 비용(Read page count)이 거의 일정하여 초기의 성능을 유지하고 있으므로 본 연구에서 제안한 단편화 방지 기법이 성능 향상에 큰 효과가 있음을 보여주고 있다.

7. 결 론

DB 기반 내비게이션은 기존의 파일 기반 내비게이션에서 제공하지 못하는 점증적 맵 갱신을 위한 새로운 접근 방법으로서 주목받고 있다.

본 논문에서는 점진적 맵 갱신을 위한 DB 기반 내비게이션에서 문제가 되는 데이터 단편화에 따른 지속적 사용에 따른 성능 저하 문제를 해결하기 위하여 새로운 MESH-ID 기반의 데이터 단편화 방지 기법을 제시하였고 이를 실제 모바일 DBMS에 적용하여 단편화 현상을 방지하여 성능 저하 문제를 해결할 수 있음을 검증하였다. 본 연구에서 제안한 기법은 점증적 맵 갱신을 위한 DB 기반 내비게이션 서비스 상용화에 기여할 수 있을 것이다.

향후, 자율주행차의 기반 기술인 고정밀맵 처리 및 갱신 기술에 적용하기 위한 추가 연구가 필요할 것이다.

References

[1] "Mobile DB solution with incremental map updates for navigation" WBS Final report, RealTimeTech Inc, 2014.
 [2] "DB Update Control Module System Based On Transaction" WBS report, RealTimeTech Inc, 2013.
 [3] NDS (Navigation Data Standard) [Internet] [https://nds-](https://nds-association.org)

[association.org](https://nds-association.org).
 [4] DERTICO [Internet] <https://europa.eu/project/actual-and-dynamic-map-transport-telematic-applications#tab-results>.
 [5] Kyoung Wook Min, Kyoung Hwan An, Ju Wan Kim, and Sung Il Jin, "The Development and Performance Evaluation of the Mobile Spatial DBMS for the Partial Map Air Update in the Navigation," *The KIPS Transactions:PartD*, Vol.15, No.5, pp.609-620, 2008.
 [6] Sung-Soo Kim, Jun-Young Baek, and Bum-Soo Kang, "Group Search Optimization Data Clustering Using Silhouette," *The Korea Operations Research and Management Science Society*, Vol.42, No.3, pp.25-34, 2017.
 [7] KairosLite (DBMS) Manual, RealTimeTech Inc.
 [8] Kyoung Wook Min, Jeong Dan Choi, and Ju Wan Kim, "Flash-aware Page Management Policy of the Mobile DBMS for Incremental Map Update," *Spatial Information Reserch*, Vol.20, No.5, pp.67-76 2012.
 [9] Kim Chang Hee, "Study on physical index-producing methods for designing and embodying databases," Graduate School of Business IT, Kookmin University, 2005.
 [10] Y. J. Joo, J. Y. Kim, and S. H. Park, "Design and Implementation of Map Databases for Telematics and Car Navigation Systems using an Embedded DBMS," *The Journal of GIS Association of Korea*, Vol.14, No.4, pp.379-389, Dec. 2006.
 [11] KyongWook Min and JeongDan, Choi, "Flash-aware Page Management Policy of the Mobile DBMS for Incremental Map Update," *IEICE Transactions on Information and Systems*, Vol.96, No.5, pp.1211-1214, 2013.
 [12] Young-Jin Joo and Hyun-Gi Kim, "A Sutdy on Unformatted Streaming Data Repository Based on Relational Indexing-Focus on Open-Source Spatial DBMS," *Journal of the Korean Institute of Plant Engineering*, Vol.18, No.3, pp.65-73, 2013.



김 용 호

<https://orcid.org/0000-0002-1828-1002>
 e-mail : kimyh@o.cnu.ac.kr
 2003년 충남대학교 전기정보통신공학부 (학사)
 2011년 ~ 현 재 충남대학교 컴퓨터공학과 석사과정

2011년 ~ 현 재 (주)리얼타임테크 내장형DBMS엔진팀 주임연구원
 관심분야 : In-Memory DBMS, Embedded DBMS



김재광

<https://orcid.org/0000-0003-0843-8990>

e-mail : jkkim@realtime-tech.co.kr

2000년 원광대학교 전자공학과(학사)

2009년 충남대학교 컴퓨터공학과(석사)

2000년 ~ 현 재 (주)리얼타임테크

내장형DBMS엔진팀 팀장

관심분야: In-Memory DBMS, Embedded DBMS



진성일

<https://orcid.org/0000-0002-1203-0333>

e-mail : sijin@cnu.ac.kr

1978년 서울대학교 계산통계학과(이학사)

1980년 KAIST 전산학과(공학석사)

1994년 KAIST 전산학과(공학박사)

1981년 ~ 현 재 충남대학교 컴퓨터공학과
교수

2000년 ~ 현 재 (주)리얼타임테크 대표이사

관심분야: In-Memory DBMS, Embedded DBMS, Spatial
DBMS