

# 생활인구 측정을 위한 가명정보의 결합키연계정보 생성방안

이 수 미\*, 신 용 태°

## A Study on Creating Linkage Keys Using Pseudonymized Data for Estimating the De Facto Population

Sumi Lee\*, Yongtae Shin°

### 요 약

본 연구는 '인구감소지역 지원 특별법'을 기반으로 생활인구의 개념을 활용하여, 대용량 데이터 연산과 결합키 연계정보 생성의 업무 부담을 감소시키고 기반 데이터의 정확성을 높이는 방안을 제안한다. 생활인구는 주민등록 인구, 체류인구, 외국인인구를 포함하며, 가명정보 결합을 통해 월별로 측정된다. 연구는 데이터의 효율적인 결합을 위해 시계열 키를 제외한 결합키 구성을 제안하고, 중복 결합키 문제에 대응하기 위한 대안법을 제시한다. 결합키연계정보의 속도와 효율성 향상을 목표로, 실험을 통해 개선된 프로세스가 기존 방식보다 효율적임을 확인한다. 이 연구는 지역별 인구 활동성의 정밀한 파악과 지방 소멸 대응을 위한 정책 수립에 중요한 기초 자료로 활용될 수 있다.

**Key Words** : de facto population, pseudonym combination information, combination key linkage information, Pseudonym information combination agency

### ABSTRACT

This study utilizes the concept of the "defacto population" based on the "Special Act for Support in Population Decline Areas" to propose methods for reducing the workload associated with large-scale data operations and the creation of linkage key information, thereby enhancing the accuracy of foundational data. The defacto population includes registered residents, transient residents, and foreign residents, and is measured monthly through the anonymized information combination. The research suggests a linkage key structure excluding time-series keys to efficiently combine data and presents alternative solutions to address issues with duplicate linkage keys. Aiming to improve the speed and efficiency of linkage key information, experiments confirm that the improved process is more effective than traditional methods. This research will be useful as fundamental data for precisely understanding regional population activity and formulating policies to counteract regional decline.

\* First Author : Soongsil University Department of IT Policy Management, leesumi@soongsil.ac.kr, 정회원

° Corresponding Author : Soongsil University Department of IT Policy Management, shin@ssu.ac.kr, 정회원

논문번호 : 202408-165-D-RN, Received August 4, 2024; Revised October 28, 2024; Accepted December 11, 2024

## I. 서 론

저출산, 고령화와 더불어 일자리 부족으로 인해 지방 중소도시의 인구 감소가 지속되고 있다. 한국고용정보원의 자료에 따르면, 전국 228개 시군구 중에서 소멸 위험이 있는 지역은 2013년 75곳에서 2022년에는 113 곳으로 늘어나고 있다. 행정안전부는 2021년 89개 시군을 인구감소지역으로 지정하고, 지역 소멸에 대응하기 위해 2022년부터 매년 1조 원 규모의 지방소멸대응기금을 이들 지역에 투자하고 있다. 또한, 2022년 '인구감소 지역 지원 특별법'에 '생활인구' 개념을 도입하여 기존의 정주인구 중심의 인구 관리 정책을 체류인구를 포함하는 개념으로 확대하고 있다.<sup>[1]</sup> 생활인구는 거주가 아닌 생활을 중심으로 인구를 바라보는 새로운 모델로, 직장, 학교, 관광, 휴양 등을 목적으로 체류하는 인구를 포함한다.<sup>[2]</sup> 이는 지역의 활력을 높이는 사람들까지 포함하여 국가 총인구 감소 상황에서 보다 현실적인 방안이며, 교통과 통신 발달로 인한 이동성 증가, 여가 중시, 일과 생활의 균형 등 현대 트렌드를 반영한다.<sup>[3]</sup> 본 연구에서는 인구감소 지역의 지원을 위해 신설된 생활인구 개념을 기반으로 명확한 선정기준과 측정을 통해 생활인구의 합리적인 활용방안을 모색하고자 한다. 생활인구 측정을 위한 활용 데이터는 행정안전부 주민등록, 법무부의국인등록, 재외동포거소신고 자료와 통신3사(SK 텔레콤, KT, LG U+)의 모바일 이동 자료를 가명결합한 데이터이다.<sup>[4]</sup> 가명데이터의 결합구조는 아래 [그림 1]과 같다.

가명정보 결합을 수행하려면, 결합기관리기관이 결합키연계정보를 생성하여 결합전문기관에 전송해야 하며, 결합전문기관은 이 정보를 바탕으로 가명정보 결합을 진행한다. 선행 연구들은 주로 인구 감소와 지역 활성화에 초점을 맞추어 생활인구를 연구해 왔으나, 가명정보 결합에 있어 결합키연계정보의 생성과 관련된 연구는 아직 미흡하다. 따라서 본 연구는 주민정보와 이동

통신정보의 결합을 위해 필요한 결합키연계정보를 효율적으로 생성하기 위해 다음과 같은 방안을 제시한다.<sup>[12]</sup>

1. 결합키연계정보 생성 시 소요되는 리소스 및 시간을 줄이는 방법을 검토한다.
2. 결합키 중복으로 인한 데이터 누락을 줄이는 방안을 제시해보고자 한다.

## II. 관련 연구

### 2.1 생활인구

생활인구는 현대 사회에서 교통과 통신의 발달로 인한 이동성과 활동성 증가를 반영하여, 2023년 1월부터 도입된 개념이다. 이 제도는 전통적인 주민등록인구뿐만 아니라, 월 1회 이상 하루 3시간 이상 특정 지역에 체류하는 사람, 그리고 외국인들까지 포함한다.

「인구감소지역 지원 특별법」 제 2조 2항에서 '생활인구'를 [그림 2]와 같이 세 가지 유형으로 정의한다.<sup>[5]</sup> 첫 번째는 '주민등록인구'로서, 「주민등록법」 제6조 제1항에 따라 주민으로 등록된 사람을 의미한다. 두 번째는 '체류인구'로서, 통근, 통학, 관광, 휴양, 업무, 또는 정기적 교류 등의 이유로 특정 지역을 월 1회 이상 방문하여 머무르는 사람을 말한다. 세 번째는 '외국인인구'로, 「출입국관리법」 제31조에 따라 외국인등록을 한 사람과 「재외동포의 출입국과 법적 지위에 관한 법률」 제6조에 따라 국내거소신고를 한 사람으로 정의하고 있다.

또한 「생활인구의 세부요건 등에 관한 규정」(행정안전부 고시)에서는, 생활인구는 체류횟수가 월 1회 이상인 사람이란 국내거소신고한 시·군·구 이외의 시·군·구에 1일 동안 머무른 시간의 총합이 3시간 이상인 경우가 월 1회 이상인 사람을 말한다.

이외에도 체류인구의 정의는 다양한 해석이 있다. 법률에서는 통근, 통학, 관광, 휴양, 업무, 정기적 교류 등 다양한 목적으로 특정 지역을 월 1회 이상 방문하여 체류하는 사람을 체류인구로 보고 있다. 그러나, 안소연 외(2023)는 이를 다르게 해석하여, 주민등록이 되어 있

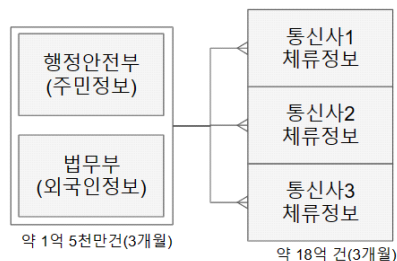


그림 1. 가명데이터 결합구조  
Fig. 1. Pseudonymized Data Combination Structure

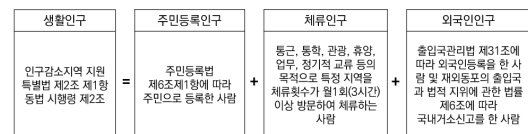


그림 2. 「인구감소지역 지원 특별법」상 '생활인구'의 개념<sup>[11]</sup>  
Fig. 2. Concept of 'Living Population' under the Special Act for Support in Population Decline Areas<sup>[11]</sup>

지 않은 상태에서 특정 지역에 1박 이상 머무르는 인구로 정의하여 법률의 정의와 차이가 있다.<sup>[6]</sup> 통계청(2020)은 유동인구를 다른 지역에서 2시간 이상 머물던 사람이 이동하여 해당 지역에서도 2시간 이상 머무르는 경우로 정의한다.<sup>[7]</sup> 한국관광공사(2022)는 방문인구를 일상생활권(거주, 통근, 통학 등)을 벗어나 관광 등의 목적으로 특정 장소에 일정 시간 이상 머무르는 사람으로 정의하며, 여기서 '일정 시간'은 해당 지자체 내의 특정 기지국에 30분 이상 체류하는 것을 의미한다.<sup>[8]</sup>

이러한 정의들을 비교해보면, 법률상의 체류인구는 단순히 잠만 방문하는 인구가 아니라, 일정 시간 이상 체류하거나 1박 이상 머무르는 사람을 포함하는 것으로 볼 수 있다. 즉, 월 1회 이상 3시간 이상 체류하거나, 최소 1박 이상 머무르는 사람들을 포함하는 개념으로 이해할 수 있다.

## 2.2 가명정보결합의 개념

가명정보결합은 크게 내부 결합과 외부 결합으로 나누어진다. 내부 결합은 같은 개인정보처리자가 보유하고 있는 개인정보를 가명 처리하여 결합하는 것이고, 외부 결합은 서로 다른 개인정보 처리자가 보유하고 있는 개인정보를 가명처리하여 결합하는 것이다. 외부 결합의 경우 제3의 기관을 통해서만 결합을 수행할 수 있다.

### 2.2.1 가명정보 결합

통계작성, 과학적 연구, 공익적 기록보존 등을 위한 목적으로 서로 다른 개인정보처리자가 보유하고 있는 데이터를 활용하고자 하는 경우 개인정보보호위원회 또는 중앙행정기관의 장이 지정한 결합전문기관을 통하여 결합이 가능하다.

가명정보 결합 절차는 [그림 3]과 같이 ① 결합신청

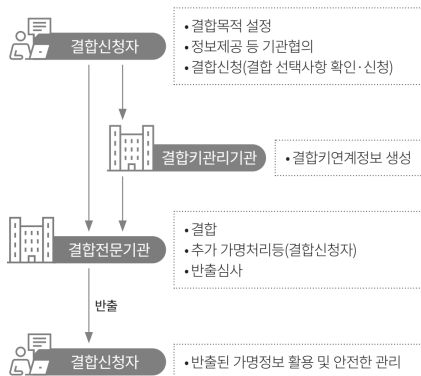


그림 3. 가명결합 프로세스<sup>[9]</sup>  
Fig. 3. Pseudonymized Data Combination Process<sup>[9]</sup>

자의 결합신청, ② 결합키관리기관<sup>1)</sup>의 결합키연계정보 생성, ③ 결합전문기관의 가명정보 결합 및 반출, ④ 결합신청자의 반출정보 활용의 순서로 진행된다.<sup>[10]</sup>

### 2.2.2 결합키 생성<sup>[9]</sup>

결합키는 일반적으로 성명, 전화번호, 생년월일 등과 같은 개인 식별 정보를 사용하여 생성된다.

[그림 4]와 같이 결합키 생성 시에는 개인을 직접 식별할 수 없도록 일방향 암호화 알고리즘을 사용하며, 가명 정보의 보안을 위해 Salt 값을 포함하여 SHA256 이상의 알고리즘을 사용하는 것이 권장된다. 여기서 Salt 값은 동일한 데이터에 대해 각 기관이 같은 해시 결과를 생성하여 데이터 결합이 가능하도록 하는 목적을 지닌다. 이를 위해 각 기관에서 동일한 Salt 값을 사용하는 것이 필요하며, 이를 통해 결합키 생성 과정에서 데이터의 일관성을 유지할 수 있다.

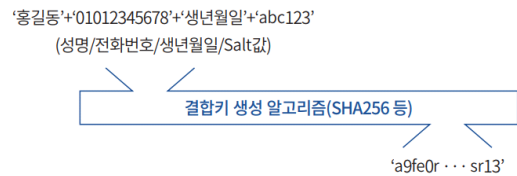


그림 4. 결합키 생성 예시<sup>[9]</sup>  
Fig. 4. Example of Linkage Key Generation<sup>[9]</sup>

### 2.2.3 결합키연계정보생성

가명정보 결합 과정은 아래 [그림 5]와 같이 결합신청자가 결합키와 일련번호를 생성하여 결합키관리기관

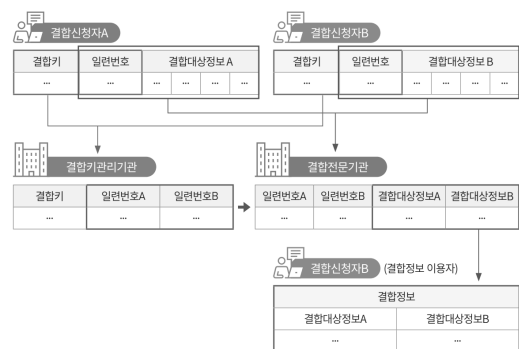


그림 5. 결합키연계정보 생성 및 결합 메커니즘<sup>[9]</sup>  
Fig. 5. Creation of Combination Key Linkage Information and Combination Mechanism<sup>[9]</sup>

1) 특정 개인을 알아볼 수 없도록 가명정보의 안전한 결합을 지원(결합키연계정보를 생성하여 결합전문기관에 제공하는 등) 하는 업무를 하는 한국인터넷진흥원 또는 개인정보위가 지정하여 고시하는 기관

에 제출한다. 결합키관리기관은 일련번호와 결합키의 매핑 정보를 생성하고, 이를 결합전문기관에 전달한다.

결합전문기관은 결합키연계정보를 기반으로 가명정보를 결합하고 결합이 완료되면, 반출심사를 거쳐 결합된 정보를 결합신청자에게 전달한다.

## 2.2.4 가명정보 결합 사례<sup>[10]</sup>

### 2.2.4.1 교통약자 보호구역 지정 분석(양평군팀, 2023)

양평군은 은퇴자의 귀촌율이 높은 지역 특성을 가지고 있어 군민의 26.7% 이상이 65세 노인이다. 현재 노인 보호구역 2곳, 어린이보호구역 26곳을 지정하였으나 대부분이 초등학교, 복지시설 등 건물 주변 위주기에 교통약자의 보행패턴 등을 반영한 보호구역 지정이 필요하여 [표 1]과 같이 양평군 주민등록정보와 통신사 정보를 결합하여 분석을 수행하였다.

가명결합 데이터를 분석 결과 양평군에 거주하고 있는 교통약자 및 유동 인구 등을 행정리 단위로 분석하고 교통사고 발생의 유형별·지역별 특성을 반영하여 [그림 6]와 같이 교통약자 보호구역 지정이 필요한 지역을 선정하였다.

표 1. 교통약자 보호구역 지정 분석을 위한 데이터<sup>[10]</sup>  
Table 1. Data for Analysis of Designated Protection Zones for Vulnerable Road Users<sup>[10]</sup>

| 보유기관    | 데이터 개요     | 레코드 수    |
|---------|------------|----------|
| 경기도 양평군 | 양평군 주민등록정보 | 약 12만건   |
| LG 유플러스 | 이동통신 정보    | 약 6천3백만건 |



그림 6. 교통약자 보호구역 필요지역 시각화<sup>[10]</sup>  
Fig. 6. Visualization of Areas Needing Designated Protection Zones for Vulnerable Road Users<sup>[10]</sup>

### 2.2.4.2 제주 한달살이 분석(카름누리팀, 2023)

코로나19로 침체되어있던 관광수요를 회복하고 관광산업을 활성화하기 위하여 데이터 활용 수요가 증가하고 있다. 개인 단위의 데이터 분석을 통하여 장기체류 관광객의 특성과 여행패턴, 관광 선호도 분석이 필요하여, [표 2]와 같이 통계청의 인구·가구·주택 통계등록정

표 2. 제주 한달살이 분석을 위한 데이터<sup>[10]</sup>  
Table 2. Data for Analysis of One-Month Stay in Jeju<sup>[10]</sup>

| 보유기관 | 데이터 개요         | 레코드 수    |
|------|----------------|----------|
| 통계청  | 인구·가구·주택통계등록정보 | 약 4천4백만건 |
| SKT  | 이동통신 정보        | 약 1천7백만건 |

보와 통신사 정보를 가명결합하였다.

분석결과 제주 한달살이 방문자의 특성은 30대 이하, 수도권 거주자, 가족 동반, 고소득자일수록, 모바일에서 레저 콘텐츠를 많이 이용할수록 제주 한달살이 경향이 크게 나타났다. [그림 7]과 같이 선호 숙박지로는 시내를 벗어난 읍·면 지역을 선호하고, 여름철에는 북쪽 해변지역이 상대적으로 높게 나타났다. 제주 한달살이 선호 방문지는 여름에는 표선, 상선 등의 선호도가 높았다.

이 외에도 가스 및 통신 데이터 결합을 통한 취약계층 생활패턴 분석, 에너지 사용량 데이터를 활용한 에너지 이용 소외계층 파악, 정책 소외계층인 청년들의 실태 조사 및 정책 개발 등 다양한 가명정보 결합이 이루어지고 있다. 본 연구는 이와 같은 가명정보를 활용하여 여러 데이터를 효율적으로 결합하는 방법에 중점을 두고 있다. 데이터의 양과 복잡성이 증가함에 따라 결합키 연계정보의 효율적인 생성과 관리가 데이터 통합의 품질에 중요한 영향을 미치며, 결합키 연계정보를 효율적으로 생성하고 활용하는 것이 필요하다.

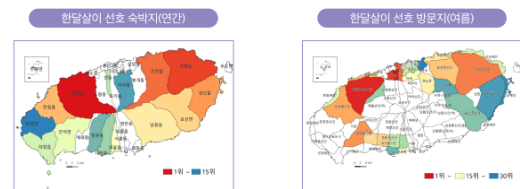


그림 7. 제주 한달살이 분석 결과 시각화<sup>[10]</sup>  
Fig. 7. Visualization of Analysis Results for One-Month Stay in Jeju

## III. 연구문제

3.1 결합키연계정보 생성 시 리소스 및 시간 소모  
생활인구 산정을 위해 가명정보 결합을 할 때, 결합키관리기관은 결합키연계정보를 생성해야 한다.

데이터 사이즈를 고려해 보면, [그림 8]과 같이 주민 정보/외국인정보의 5천만건의 3개월 데이터는 약 12GB이고, 통신사 데이터는 3개월 데이터는 약 144GB로, 이 많은 양의 데이터가 결합키관리기관으로 전송되어야 한다.

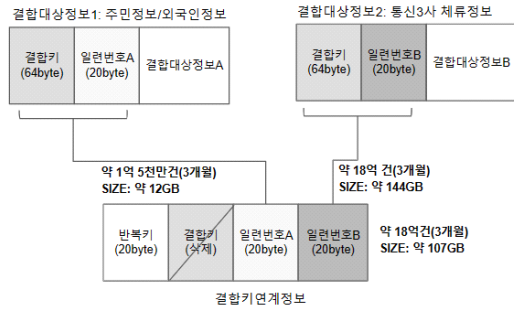


그림 8. 결합키/일련번호 사이즈 및 결합키연계정보 구성  
Fig. 8. Combination Key/Serial Number Size and Structure of Combination Key Linkage Information

결합의뢰기관의 결합키/일련번호의 사이즈가 커서 [그림 9]와 같이 결합키관리기관으로 여러 조각으로 나누어 전송하였고, 각각 결합키연계정보를 생성하였다.

이렇게 나누어진 데이터를 처리하기 위해, 결합의뢰기관이 결합키/일련번호 전송하고, 결합키관리기관은 결합키연계정보 생성하며 이를 결합전문기관 전송하고, 결합전문기관이 데이터를 결합하여 반출하는 과정을 반복적으로 수행해야 한다. 이러한 반복적인 프로세스는 상당한 시간과 리소스가 소모된다. 데이터 결합의 복잡성과 리소스 소모를 줄이기 위해 결합키연계정보 생성 방법의 개선이 필요하다.

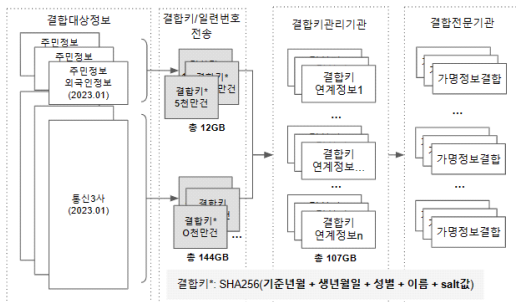


그림 9. 분할 생성된 결합키연계정보를 이용한 가명정보결합  
Fig. 9. Pseudonymized Data Combination Using Segmented Generated Linkage Key Information

### 3.2 중복 결합키 삭제로 인한 생활인구 데이터 누락

데이터 결합 과정에서 결합키는 두 데이터셋을 연계하기 위한 고유한 식별자로 사용된다. 그러나 결합키를 생성할 때, 입력되는 데이터가 같으면 해시 결과도 동일하기 때문에 중복이 발생할 수 있다. NICE신용평가정보에 따르면, 주민등록상 생년월일이 일치하는 사람 중 같은 이름을 가진 경우가 11.7%라는 보고가 있다.<sup>[11]</sup>

이는 결합키 생성 시 같은 생년월일과 이름을 가진 사람들 사이에서 중복이 발생할 가능성이 높다는 것을 의미한다.

행정안전부의 주민정보 결합키 생성 과정에서도 이러한 중복 현상이 나타난다. 위의 추정에 따르면, 데이터 중복으로 인해 생성된 결합키 중 약 10% 정도가 중복될 수 있다. 이로 인해 결합 후 데이터가 의도와 다르게 해석될 수 있으며, 중복 결합키를 삭제하면 약 500만 명의 정주 인구 데이터가 누락되게 된다. 이는 인구 정책을 수립하는 데 있어 중요한 데이터의 손실을 야기할 수 있으므로, 정확한 인구 통계 데이터를 확보하기 위해서는 이러한 문제에 대한 보완이 필요하다.

## IV. 개선 방안

### 4.1 결합키연계정보를 위한 대상데이터 사이즈 축소

결합의뢰기관이 결합키관리기관에 전송할 결합키/일련번호 쌍을 구성할 때, 중복을 제거하고 유일한 정보만으로 구성한다. 주민등록정보 및 외국인정보의 3개월치 데이터 중 유일한 정보만을 사용하고, 통신 3사 역시 유일한 결합키/일련번호 쌍을 생성하여 결합키관리기관에 전송하는 한다. 이렇게 하면 결합키 연계 정보를 생성하는 기반 데이터의 크기가 크게 줄어들게 된다.

결합키를 구성할 때 [표 3]과 같이 기준년도를 결합키에서 제외한 것은 시계열 정보를 제외하고 매핑 마스터로 정의하여 결합키연계정보의 크기를 줄일 수 있다. 이 방식은 결합키 연계정보 생성의 효율성을 높이기 위한 접근으로, 시계열 요소를 포함하지 않고도 효과적으로 매칭이 가능한 구조를 만들기 위한 것입니다. 다만, 가명정보 결합 시 기준년월 별로 나누어 결합해야 하는 제약이 따른다.

구체적으로, 아래 [그림 10]와 같이 주민정보 및 외국인정보의 경우 결합대상정보의 크기가 3개월에서 1개월로 약 1/3 수준으로 줄어들고, 통신 3사의 데이터는 1/30 정도로 감소한다. 이를 통해 결합키관리기관은 기반 데이터의 규모가 줄어들고, 결합키연계정보를 한 번만 생성해도 되는 등 효율성을 크게 향상시킬 수 있다. 이러한 개선은 가명정보 결합 과정에서 시간과 리소스를 절약할 수 있어 전체적인 작업 효율을 높일 수 있다.

표 3. 결합키 구성 방법 제안  
Table 3. Proposed Method for Linkage Key Configuration

|    |                              |
|----|------------------------------|
| 기존 | SHA256(기준년월+생년월일+성별+이름+salt) |
| 개선 | SHA256(생년월일+성별+이름+salt)      |

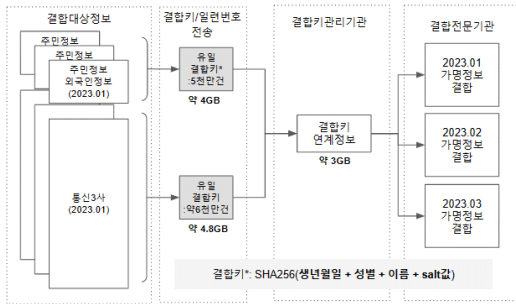


그림 10. 개선된 결합키연계정보 생성 방안  
Fig. 10. Proposed Method for Improved Linkage Key Information Generation

#### 4.2 중복 결합키를 결합키연계정보에 추가

중복된 결합키를 삭제하고 결합하는 과정에서 데이터의 손실이 발생할 수 있다. 이를 해결하기 위해 제안할 수 있는 방법은 ① 중복 결합키가 발생하는 정보를 분리한다. ② 유일한 결합키가 있는 주민정보를 기준으로 체류정보(통신)를 join하여 결합키연계정보 생성한다. ③④ 분리했던 중복 결합키의 일련번호들을 결합키연계정보에 추가한다. 이는 [그림 11]과 같다. 이와같이 생성된 결합키연계정보를 기반으로 결합을 하게 되면 주민정보와 체류정보에서 결합키를 생성할 때 중복되는 정보를 누락하지 않고 결합을 할 수 있어 데이터의 손실을 줄일 수 있다.

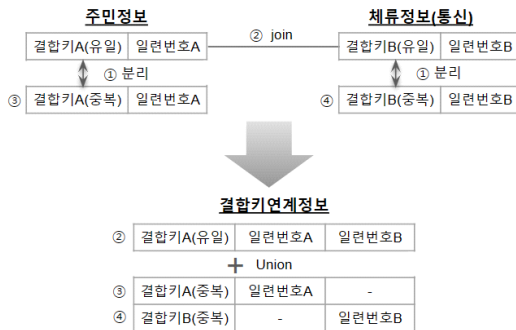


그림 11. 결합키연계정보에 중복 결합키의 일련번호 정보 추가  
Fig. 11. Adding Serial Number Information of Duplicate Linkage Keys to Linkage Key Information

### V. 실험연구

이 장에서는 기존 결합키연계정보 생성 속도를 측정하고, 결합키연계정보 생성을 위한 대상 데이터의 사이즈를 축소하여 얼마나 단축되는지를 실험하였다.

#### 5.1 환경구성

본 실험은 대규모 데이터(약 18억 건)의 처리를 위해 고성능 Windows 서버(700GB 메모리) 환경에서 Jupyter Notebook과 Python을 사용하여 결합키/일련번호 데이터 세트를 생성하였다. 대용량 데이터의 성능 최적화를 위해 고사양 환경에서 실험을 진행하였으며, 이는 실험실 조건에서 연구의 타당성을 검증하기 위함이다.

#### 5.2 데이터세트 구성

파이썬을 활용하여 3개월치의 데이터인 주민정보 1억 5천 건, 통신정보 18억 건의 결합키/일련번호 세트를 생성하였다. 개선안에 사용될 결합키/일련번호 세트를 5천만건과 6천만건씩 생성하여 테스트에 사용하였다. 실험에서 사용한 스크립트는 [그림 12]과 같다.

```
import pandas as pd
import hashlib

data = []
for i in range(50000000):
    original_string = str(i) # 임의의 문자열
    encoded_string = original_string.encode() # 문자열을 바이트로 인코딩
    sha256_hash = hashlib.sha256(encoded_string).hexdigest() # SHA256 해시 생성
    data.append({
        'key': sha256_hash,
        'df_A': original_string
    })
df = pd.DataFrame(data)
```

그림 12. 가상 데이터 생성스크립트  
Fig. 12. Virtual Data Generation Script

#### 5.3 결합키연계정보 생성단계별 성능측정 결과

결합키연계정보를 생성하기 위한 단계로 업로드, 매핑정보생성, 다운로드의 총 3단계로 정의하였다. 업로드 단계는 생성된 테스트 데이터를 결합키를 위해 시스템에 업로드하는 시간을 의미한다. 매핑정보 생성단계는 결합키관리기관에서 결합키 연계 정보를 생성하는 단계이다. 다운로드 단계는 결합키연계정보를 결합전문기관에 전달하기 위하여, 생성된 매핑정보를 다운로드하는 단계이다.

구성된 데이터 세트로 진행한 각 단계별 성능 측정 결과는 [표 4]와 같다. 데이터 업로드, 매핑정보 생성, 다운로드 각각에서 개선된 것을 확인할 수 있었다. 업로드 단계에서 약 1/18, 다운로드 단계에서 약 1/27정도로 시간이 줄어들었다. 전체적인 처리 시간을 비교했을 때, 개선 방식은 기존 방식에 비해 약 1/22 정도 절감되었다. 시계열 정보를 제외하여 키를 구성하게 되면 결합키연계정보를 생성하는 속도가 개선되는 것을 확인할 수 있었다. 이는 결합키관리기관이 결합키연계정보를 생성 작업에 소요되는 리소스와 시간이 크게 줄어들며, 전체 결합 프로세스의 효율성 향상에 기여할 수 있다.

표 4. 결합키연계정보 생성 속도 비교

Table 4. Comparison of Combination Key Linkage Information Generation Speeds

| 단계          |          | 기준(시:분:초) |         | 개선(시:분:초) |          | 비율   |
|-------------|----------|-----------|---------|-----------|----------|------|
| 업<br>로<br>드 |          | 00:47:43  |         | 00:02:39  |          | 1/18 |
|             | 주민<br>정보 | 1.5억      | 0:03:37 | 5천만       | 00:01:12 |      |
|             | 통신<br>정보 | 18억       | 0:44:06 | 6천만       | 00:01:27 |      |
|             |          |           |         |           |          |      |
| 매핑정보<br>생성  |          | 00:35:42  |         | 00:02:23  |          | 1/15 |
| 다운로드        |          | 02:13:08  |         | 00:05:00  |          | 1/27 |
| 총합          |          | 03:36:33  |         | 00:10:02  |          | 1/22 |

## VI. 결 론

본 연구는 생활인구 산정 과정에서 가명정보 결합 효율성 저하 및 중복 결합키로 인한 데이터 손실 문제를 해결하기 위한 방법을 제시하였다.

첫째, 결합키 생성 시 시계열 정보를 제외하여 키를 구성하면 결합키연계정보의 크기와 생성에 필요한 자원을 줄일 수 있다. 현재 가명정보 처리 가이드라인에는 결합키연계정보가 결합키/일련번호 매핑을 기준으로 생성되는데, 결합키에 시계열 정보를 포함하는 경우 데이터 양이 증가하면서 결합키연계정보의 크기와 생성에 필요한 자원이 증가하게 된다. 시계열 정보를 다루는 결합키의 경우 결합키연계정보를 한 번 생성하고 반복적으로 활용할 수 있도록 함으로써, 가명결합 과정의 효율성을 개선할 수 있다. 이러한 내용이 향후 가이드라인에 포함된다면 가명정보 결합이 필요한 연구자들에게 유용한 참고자료가 될 것이다.

둘째, 결합키 중복 문제를 해결하기 위한 방법은 주민등록번호를 결합키로 사용하는 것이 가장 효과적이지만, 이는 현재 법적 제한이 있어 결합키를 생성하는데 어려움이 있다. 이에 대안적 방법으로 중복 결합키를 결합키연계정보에 포함시키는 방법을 제안하고, 이러한 접근은 데이터의 품질 개선 및 생활인구 산정의 정확성을 높이는데 기여할 것으로 생각된다.

끝으로 본 연구에서 제안하는 해결 방안이 가명정보 결합에 관한 정책을 수립하고 실행하는 데 있어 중요한 자료로 활용되길 기대한다.

## References

[1] S. S. Lee, B. H. Yun, M. H. Lee, and Y. H. Kwon, "Consideration of issues for calculating

de facto population and utilizing urban policies," *J. Urban Studies and Real Estate*, vol. 14, no. 4, pp. 69-87, 2023.

(<https://doi.org/10.21447/jusre.2023.14.4.4>)

- [2] C. W. Seo and J. A. Bae, "What are the drivers affecting the formation of the living population? : Focusing on the effects of the local culture and tourism financial expenditure," *The Korea Local Administration Rev. (Krila)*, vol. 37, no. 4, pp. 222-240, 2023.  
(<http://dx.doi.org/10.22783/krila.2023.37.4.221>)
- [3] B. K. Min and J. S. Choi, "Characterizing regions based on the de facto population: Focusing on the local areas in South Korea," *J. Korean Urban Manag. Assoc.*, vol. 36, no. 4, pp. 41-60, 2023.  
(<https://doi.org/10.36700/KRUMA.2023.12.36.4.41>)
- [4] Statistics Korea and Ministry of the Interior and Safety, "Results of the De Facto population estimation in pilot areas based on pseudonymized data combination between public and private sectors," Jan. 2024.
- [5] Legislation and Law Information Center, "Special Act for Support in Population Decline Areas," 2023.
- [6] S. H. Ahn, S. J. Lee, S. H. Min, M. A. Kim, B. K. Jeon, and M. S. Kang, "The necessity of introducing transient population in the era of population decline and policy measures," *National Land Policy Brief*, pp. 1-8, 2023.
- [7] Statistics Korea, "Mobile Telecommunications Big Data-based Floating Population Mapping Service," 2020, Retrieved May 12, 2024, from <https://giraf.sktelecom.com/web/kostat/>.
- [8] Korea Tourism Organization, "Korean tourism through data," 2022.
- [9] Personal Information Protection Commission, "Guidelines for the processing of pseudonymized information," Apr. 2022.
- [10] Personal Information Protection Commission, "2023 best practices of utilizing pseudonymized information," 2024.
- [11] Yonhap News Agency, "According to NICE Credit Information, there is an 11.7% proba-

bility that individuals with the same date of birth and name exist in the resident registration records,” Sep. 17, 2012, Retrieved May 15, 2024, from <https://www.yna.co.kr/view/AKR20120914199900002>

- [12] S. M. Lee, J. S. Lee, and E. J. Kim. “A study on the generation method of combination key linkage information of pseudonym information for estimating de facto population,” in *Proc. KICS Conf.*, pp. 868-870, Jeju Island, Korea, Jun. 2024.

신 용 태 (Yongtae Shin)



1985년 2월 : 한양대학교 산업  
공학과 졸업  
1990년 5월 : University of  
Iowa 컴퓨터학 석사  
1994년 5월 : University of  
Iowa 컴퓨터학 박사  
1995년 3월~현재 : 숭실대학교  
컴퓨터학부 교수

<관심분야> 정보보호, IoT, 클라우드 컴퓨팅  
[ORCID:0000-0002-1199-1845]

이 수 미 (Sumi Lee)



2002년 2월 : 홍익대학교 컴퓨  
터공학과 졸업  
2022년 2월 : 연세대학교 공학  
대학원 인공지능 석사 졸업  
2023년 3월~현재 : 숭실대학교  
IT정책경영학과 박사과정  
<관심분야> 가명정보, 개인정  
보보안, AI/빅데이터

[ORCID:0009-0002-4521-542X]