



저작자표시-비영리-변경금지 2.0 대한민국

이용자는 아래의 조건을 따르는 경우에 한하여 자유롭게

- 이 저작물을 복제, 배포, 전송, 전시, 공연 및 방송할 수 있습니다.

다음과 같은 조건을 따라야 합니다:



저작자표시. 귀하는 원저작자를 표시하여야 합니다.



비영리. 귀하는 이 저작물을 영리 목적으로 이용할 수 없습니다.



변경금지. 귀하는 이 저작물을 개작, 변형 또는 가공할 수 없습니다.

- 귀하는, 이 저작물의 재이용이나 배포의 경우, 이 저작물에 적용된 이용허락조건을 명확하게 나타내어야 합니다.
- 저작권자로부터 별도의 허가를 받으면 이러한 조건들은 적용되지 않습니다.

저작권법에 따른 이용자의 권리는 위의 내용에 의하여 영향을 받지 않습니다.

이것은 [이용허락규약\(Legal Code\)](#)을 이해하기 쉽게 요약한 것입니다.

[Disclaimer](#)

정 재 준 교수지도
석사학위 청구논문

새주소체계 활용을 위한
지오코딩 적용 방안

2009

성신여자대학교 대학원
지리학과
노혜정

새주소체계 활용을 위한
지오코딩 적용 방안

정 재 준 교수지도

이 논문을 석사학위논문으로 제출함.

2008년 11월

성신여자대학교 대학원
지리학과
노혜정

인준서

노혜정의 석사학위 논문으로 인준함.

심사위원 _____ 인

심사위원 _____ 인

심사위원 _____ 인

성신여자대학교 대학원

논문개요

주소는 현대사회에서 사람들이 위치를 개념화하는 기본적인 수단 중 하나이며 사람들은 위치를 점유하고 있다. 이렇게 사람들이 위치하고 있는 곳의 정보를 지니는 주소는 개인행태나 특정사상의 공간적 분포 분석 시 중요한 정보를 지닌다.

한편, 현재 우리나라의 주소제도는 지번체계에서 도로망 체계로 변화하고 있다. 기존의 지번주소가 급속한 산업화와 인구의 증가 및 도시의 팽창 등으로 인해 위치정보 제공이라는 본연의 목적을 상실하였기 때문이다. 이에 따라 도로를 중심으로 건물번호를 체계적으로 부여하는 새로운 주소체계를 도입하게 되었으며, 2008년 현재 지번체계와 함께 병행하여 사용하고 있다.

지오코딩(geocoding)은 경위도나 직각좌표와 같은 지오코드(geocode)를 할당하는 과정으로, 본 연구에서의 지오코딩은 주소정합의 의미로 주소 내 행정구역 및 지번과 같이 특수한 요소들을 통해 x, y 좌표를 할당받는 과정이다. 주소와 같이 문자로 된 위치정보가 지오코딩되면 지리정보시스템(GIS, Geographic Information System)에서 공간 분석 및 모델링에 사용할 수 있다.

새주소체계 변경 전 지오코딩 연구는 필지기반 지오코딩이 사용되어 왔지만, 필지번 DB 등 데이터 접근성이 낮아 활용도가 적었다. 한편, 우리나라의 주소체계 변화는 새주소체계가 도로망주소이므로, 도로망기반 지오코딩을 활용할 수 있으며 도로 세그먼트 정보는 인터넷에서 얻을 수 있는 장점을 가지고 있다.

본 연구는 이러한 주소체계의 변화에 따라 도로망기반의 지오코딩의 적용가능성을 확인하고, 지오코딩의 품질개선을 위한 방안을 제시하고자 하였다.

본 연구를 통해 얻은 결론은 다음과 같다.

첫째, 새주소체계의 변화에 따라 기존에 적용할 수 없었던 도로망기반의 주소체계를 이용한 지오코딩이 가능해졌다. 둘째, 주소를 특징에 따라 세분화하고 이에 대한 도로 세그먼트를 구축하여 지오코딩을 실시하면 정합률을 향상시킬 수 있다. 셋째, 도로 세그먼트 구축 시 주소의 범위 및 특징에 따라 도로중심선을 수정함으로써 위치정확도를 증가시킬 수 있다.

본 연구는 주소체계 변화에 따라 새주소체계 활용하기 위한 지오코딩 적용가능성과 품질개선 방안에 대해 살펴보았으며, 지오코딩에 대한 기초적인 연구로서 향후 점 사상의 데이터 구축에 효과적으로 사용할 수 있을 것으로 생각된다.

목 차

제1장 서론	1
제1절 연구의 배경 및 목적	1
제2절 연구동향	2
1. 지오코딩 관련 연구	2
2. 새주소체계 관련 연구	4
제3절 연구의 내용 및 방법	5
제2장 이론적 배경	7
제1절 지오코딩 이론 및 모델	7
1. 정의 및 개념	7
2. 지오코딩 프로세스	9
3. 주소모델	10
4. 품질평가	14
제2절 주소체계	16
1. 주소표시방법	16
2. 우리나라 주소체계	18
제3장 지오코딩 적용 실험	29
제1절 실험 계획 및 연구지역 설정	29

제2절 데이터 구축 및 지오코딩	32
1. 도로 세그먼트	32
2. 참조자료	34
3. 주소데이터 수집 및 변환	35
4. 지오코딩	37
제3절 품질평가	41
1. 정합률	41
2. 위치정확도	41
제4장 지오코딩 품질 개선 실험	45
제1절 정합률 개선 실험	45
1. 가지번호(전체)	45
2. 가지번호(일부)	49
제2절 위치정확도 개선 실험	50
1. 기초번호 및 가지번호2	50
2. 가지번호1	55
제3절 결과 분석	59
1. 정합률	59
2. 위치정확도	59
제5장 결론	62
참 고 문 헌	
ABSTRACT	

표 목 차

[표 1] 주소표시방법의 장·단점	18
[표 2] 새주소 표기방법(예)	23
[표 3] 기초번호의 도로 세그먼트	33
[표 4] 주소표시 형식 변환(예)	36
[표 5] ArcGIS Street map에서 제공하는 address locator type	38
[표 6] ArcGIS의 지오코딩 옵션	39
[표 7] 전체주소 : 정합률	42
[표 8] 전체주소 : 기초번호의 위치오차 통계량	42
[표 9] 가지번호의 도로 세그먼트	46
[표 10] 가지번호 : 정합률	47
[표 11] 가지번호2 : 정합률	49
[표 12] 기초번호 : 위치오차 통계량	52
[표 13] 가지번호2 : 위치오차 통계량	53
[표 14] 가지번호1 : 위치오차 통계량	58
[표 15] 주소형태에 따른 정합률 비교	59
[표 16] 전체 주소 : 도로 세그먼트 간 위치오차 통계량 비교	60

그림 목 차

〈그림 1〉 연구 흐름도.....	6
〈그림 2〉 주소 지오코딩 개념.....	8
〈그림 3〉 web 2.0 기반 지오코딩 개념	8
〈그림 4〉 도로망모델 : 동질적	11
〈그림 5〉 도로망모델 : 이질적.....	12
〈그림 6〉 필지기반 모델	13
〈그림 7〉 address-point 모델	14
〈그림 8〉 지오코딩 품질 저하원인 및 개선방법.....	15
〈그림 9〉 지번 설정방식 : 사행식.....	21
〈그림 10〉 지번 설정방식 : 기우식	21
〈그림 11〉 지번주소의 문제점1 : 지번배열 순서가 불합리하여 순차적 위치찾 기가 어려운 경우.....	24
〈그림 12〉 지번주소의 문제점2 : 한 건물이 여러 지번에 걸쳐있는 경우.....	24
〈그림 13〉 도로명 주소 부여 순서	25
〈그림 14〉 새주소체계의 건물부여 방식.....	27
〈그림 15〉 실험지역 행정구역	29
〈그림 16〉 실험흐름도.....	31
〈그림 17〉 실험지역 도로 세그먼트.....	34
〈그림 18〉 참조자료 : 새주소안내지도 및 지적도.....	35
〈그림 19〉 address locator의 역할.....	37

〈그림 20〉 ArcGIS Street map의 매칭 알고리즘	40
〈그림 21〉 위치정확도 측정방법	43
〈그림 22〉 전체 주소 : 위치정확도	43
〈그림 23〉 전체 주소 : 위치정확도 분석(성신9길 부근)	44
〈그림 24〉 전체 주소 : 위치정확도 분석(동선길 부근 교차로)	44
〈그림 25〉 가지번호 중 가지번호 도로 세그먼트로 매칭되지 않는 경우	48
〈그림 26〉 기초번호 및 가지번호2 : 도로 세그먼트 비교	51
〈그림 27〉 기초번호 : 위치정확도	54
〈그림 28〉 가지번호2 : 위치정확도	54
〈그림 29〉 가지번호1 : 도로중심선에 기초한 도로 세그먼트 구축 예	56
〈그림 30〉 가지번호1 : 도로중심선을 수정한 도로 세그먼트 구축 예	56
〈그림 31〉 가지번호1 : 도로 세그먼트 비교	57
〈그림 32〉 도로 세그먼트 비교(전체 도로망)	57
〈그림 33〉 기초번호1 : 위치정확도	58
〈그림 34〉 전체주소 위치정확도 비교	61
〈그림 35〉 전체 주소 : 수정 도로 세그먼트의 위치정확도	61

제1장 서론

제1절 연구의 배경 및 목적

주소는 현대사회에서 사람들이 위치를 개념화하는 기본적인 수단 중 하나이며, 위치의 정보를 지니는 주소는 개인행태나 특정 사상의 공간적 분포 분석 시 사용되고 있다.

우리나라는 1910년부터 지번중심의 주소를 사용해왔다. 그러나 지번제도가 1910년 일제의 토지조사사업으로 부여된 후, 급속한 산업화와 인구의 증가 및 도시의 팽창과 각종 지역개발사업으로 인해 현재의 지번체계 상의 위치정보는 매우 복잡하여 일관성이 결여되어 있는 문제점이 지적되었다. 이에 따라 1996년부터 시범사업 등을 거쳐 도로를 중심으로 건물번호를 체계적으로 부여하는 새로운 주소체계를 도입하게 되었으며, 2008년 현재 지번체계와 병행하여 사용하고 있다.

이러한 주소체계의 변화는 도로망기반의 지오코딩(geocoding)을 가능하게 하였다. 지오코딩은 문자로 된 데이터를 이용해 경위도나 직각좌표를 할당하는 과정을 의미한다. 주소와 같이 문자로 된 위치정보가 지오코딩되면 지리정보시스템(GIS, Geographic Information System)에서 할당된 좌표를 통해 두 점의 거리를 계산, 데이터의 공간적 분포를 시각화, 센서스 트랙과 같은 어떤 구역에 자료를 할당하는 것이 가능하다(Clapp 등, 1997). 또한 지오코딩은 응급구조, 운송서비스 뿐만 아니라 범죄 지도화(mapping) 및 분석, 공공시설입지 모델링과 같은 분야에서도 널리 사용되고 있다.

우리나라의 주소체계는 약 100년 동안 지번체계이었기 때문에 지오코딩에 대한 연구 및 활용연구도 필지기반 지오코딩을 사용해 연구가 진행되어 왔다. 필지기반 지오코딩을 위해서는 해당지역에 대한 지적도와 지번 DB가 필요하지만, 이 데이터들은 데이터접근성이 낮아 결과적으로 지오코딩의 활용도를 적게 만드는 요소로 작용하고 있다.

한편, 현재 우리나라의 주소체계는 지번주소와 도로명주소가 병행되어 사용되고 있다. 새주소체계인 도로명주소는 지오코딩 시 도로망기반 지오코딩을 사용할 수 있으며, 이를 위한 각 도로의 건물범위에 대한 정보는 새주소안내시스템 (<http://www.juso.go.kr>)을 통해 얻을 수 있다. 즉, 주소체계 변화는 기존의 필지기반 지오코딩에 비해 데이터의 취득이 용이해졌음을 의미한다.

본 연구는 새주소체계인 도로명주소기반의 지오코딩의 가능성을 확인하고, 품질평가를 실시함으로써 이를 개선할 수 있는 품질개선 방안을 제시하고자 하였다.

제2절 연구동향

1. 지오코딩 관련 연구

지오코딩에 관한 연구는 지리학뿐만 아니라 보건, 컴퓨터, 정치학 등 다양한 연구 분야와 응용분야에서 진행되고 있으며 크게 지오코딩 이론 및 품질개선에 관한 연구와 이에 대한 활용연구로 나눌 수 있다.

지오코딩의 이론 및 개선에 관한 연구는 지오코딩의 유용성, 신뢰성, 정확도

를 개선하기 위한 연구들이 주를 이루고 있다. 먼저 이론적 연구로는 지오코딩 시 내삽 프로세스의 가정에 대한 연구(Dearwent 등, 2001; Karimi 등, 2004)와 장소 단위의 지오코딩의 선택에 대한 연구(Krieger, 1992; Geronimus 등, 1995; Krieger 등, 2003)가 있다. 지오코딩의 정확도에 관한 연구는 참조데이터의 정확도에 대한 연구와 참조데이터와 지오코딩 간 차이를 줄이는 방법에 대한 연구가 있는데, 참조데이터의 정확도에 관한 연구(Gatrell, 1989; Block, 1995; Drummond, 1995; Chung 등, 2004)는 자세한 공간모델에 기반을 두어 지오코딩의 오차를 줄이고자 하였다. Karimi 등(2004)과 Wu 등(2005)은 실제 데이터와 지오코딩된 feature 간 차이를 줄이는 방법에 대해 연구하였다.

이 밖에 Zimmerman 등(2007)은 거주지의 주소 지오코딩 때 발생하는 위치 오차의 확률적 분포를 모델링하였고, Yang 등(2004)은 지오코딩 시 주로 사용되는 ArcView, Automatch¹⁾, ZP4(Zip+4-level census data)²⁾의 장단점을 비교분석하였다.

지오코딩 활용 연구는 주로 보건 분야에서 이루어지고 있으며, 고객 관리 및 고객 분석에서도 활용되고 있다. Vine 등(1998)과 Boulos(2004), Rushton

1) Matchware社의 소프트웨어로써 1985년에 Matt Jaro에 의해 개발되었다. 현재 생산되지 않으나, 데이터통합시스템(data integrity solution systems)에서 광범위하게 통합되어 사용된다.

2) Semaphore社의 소프트웨어로써 데이터의 파싱(parsing) 및 표준화기능을 제공한다. 위의 논문에서는 ZP4와 zip+4-level census data를 함께 사용하였다. zip+4-level census data는 우편번호에 대한 centroid와 census tract 및 block number가 저장되어 있어 두 개의 소프트웨어를 사용하여 지오코딩을 실시하였다.

등(2006)은 보건 분야에서 지오코딩을 사용하여 연구를 진행하였으며, Olligschlaeger(1998)와 Ratcliffe(2001)는 범죄 분석에 지오코딩을 사용하였다. 이 밖에 정치학과 컴퓨터공학에서도 지오코딩을 활용한 연구가 진행되고 있다(Goldberg 등, 2007). Robert(1999)는 사회적 맥락에서의 보건효과를 추정하였고, Evans와 Kantrowitz(2002)는 지오코딩을 이용한 질병의 지리적 클러스터를 측정하였다. Kravets와 Hadden(2007)은 보건자료의 지역특성의 영향을 측정하는 연구에서 농촌과 도시지역에 대해 지오코딩된 주소와 실제위치를 비교하였다.

국내에서는 이지영과 김학열(2006)이 우리나라의 계층적 주소체계를 위한 지오코딩 방법 구현에 대한 연구가 있다. 이 연구는 지번주소체계에 적합한 지오코딩방법을 이론적으로 개발한 것으로, 2차원 및 3차원 주소정합기술로 구성된 3차원 주소위치확인에 적용할 수 있는 새로운 접근방법을 제시하였다. 손철(2002)은 지오코딩을 이용해 주택시장에 대한 특성감안함수를 추정하고, 다양한 지오코딩 방법이 추정결과에 미치는 영향을 비교함으로써 비교연구의 필요성을 제시하였다.

2. 새주소체계 관련 연구

새주소체계에 관한 연구는 1996년 시범사업이 시작된 이후 지속적으로 연구되어 왔으며 주요한 연구주제는 도로명과 건물번호에 기초한 새주소체계의 개선방안과 안내 및 관리시스템 개발이다.

개선방안에 대한 연구를 먼저 살펴보면, 이진하(2008)는 지번에 기초한 구(舊)주소와 새주소의 문제점을 파악하여 외국의 주소체계와의 비교 분석 후, 우리나라 실정에 맞는 지번과 새주소의 발전방안을 제시하였으며, 이기환(2004)은 춘천시의 사례를 바탕으로 새주소 부여사업의 유지관리 개선방안을 제시하였다. 임재일(2001)은 부산시를 사례로 간선도로망의 선정에 따른 주소체계의 개선방향과 발전방향을 제시하였고, 유상철(2008)은 도로명주소의 인지도 향상 및 조기정착을 위한 개선방향을 제시하였다.

새주소 안내 및 관리시스템 개발에 대한 연구는 특정 연구지역을 중심으로 시스템을 구축하였다. 연상호와 유준상(2001)은 제천시를 사례로 새주소체계를 기반으로 웹(web)에서 제천시 새주소 안내 시스템을 구축하였으며 이민영 등(2001)은 청주시를 사례로 도로명 및 건물부여체계를 통해 새주소관리시스템을 구현하고 평가하였다.

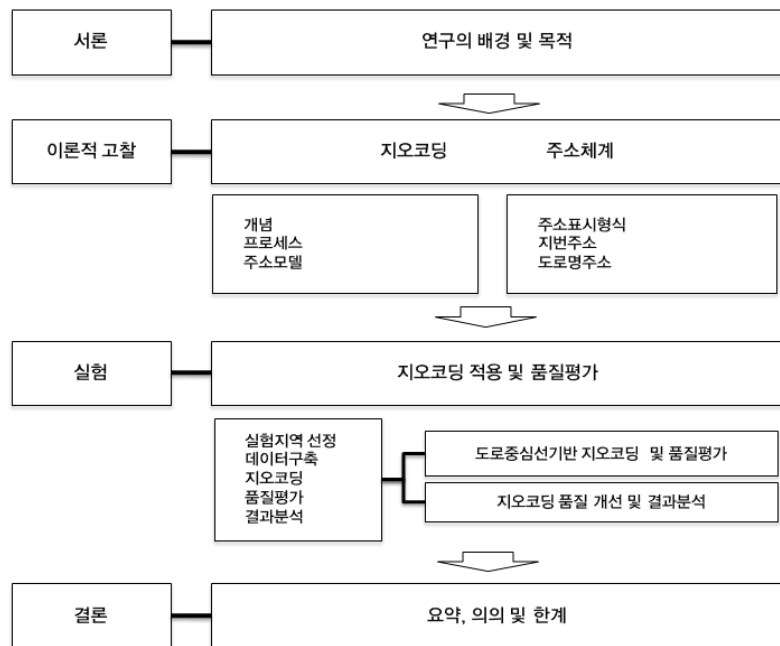
김길웅(2008)은 새주소 부여 사업의 유지관리를 효율적으로 처리하기 위한 mobile GIS 시스템을 구현하였으며, 구은훈(2005)은 경상북도 경산시의 새주소 체계 Web-GIS의 개발환경 및 시스템 구현을 통한 문제점과 이를 기반으로 개선방향 및 활용분야를 제시하였다.

제3절 연구의 내용 및 방법

본 연구는 새주소체제로 지오코딩을 하고 이에 대한 품질개선 방안을 제시하는 것을 목적으로 하고 있다. 이를 위해 본 연구에서는 서울특별시 성북구의 동선동 일대를 대상으로 도로망기반의 지오코딩을 적용하여 지오코딩의 가능성을 제시하고, 지오코딩의 품질개선 방안을 제시하였다. 또한 품질개선의 정도를

판단하기 위해 정합률과 위치정확도에 대해 기초통계량을 측정하였다. 본 연구의 전체적인 내용 및 연구방법은 다음과 같다(그림 1).

이론적 배경으로 지오코딩의 개념과 프로세스 및 주소모델과 새주소체계와 지번주소체계의 주소할당 방법을 고찰하고 새주소체계의 도로망 체계를 중점적으로 살펴봄으로써 지오코딩 시 고려해야할 측면을 살펴보았다. 위의 이론적 배경을 바탕으로 실험지역인 서울특별시 성북구 동선동 일대에 대해 지오코딩하였으며, 품질평가로써 정합률과 지적도 간의 거리차를 분석하였다. 도로중심선을 기초로 한 지오코딩 결과를 통해 이를 개선시킬 수 있는 방안들을 제시하였다. S/W로는 ESRI의 ArcGIS Street map과 Microsoft의 Access, Excel 2007을 사용하였다.



〈그림 1〉 연구 흐름도

제2장 이론적 배경

제1절 지오코딩 이론 및 모델

1. 정의 및 개념

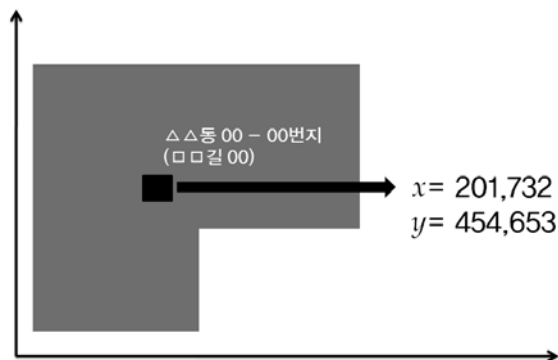
지오코딩(geocoding)은 경위도나 직각좌표와 같은 지오코드(geocode)를 할당하는 것을 의미한다. 주소 내 행정구역 및 지번과 같이 특수한 요소(location-specific element)들을 참조데이터와 비교함으로써 해당 장소의 x, y 좌표로 할당하는 과정이다(Paul, 2008). 주소와 같이 문자로 된 지리적 단어를 지오코드로 변환시켜주는 프로세스로써 Drummond(1995)와 Vine 등(1998)과 Bonner 등(2003)은 주소정합(address matching)과 동의어로 사용하고 있다(그림 2).

또한 최근 web 2.0의 논의가 진행되면서, 지오코딩의 개념은 Web 2.0 환경에서는 웹페이지, 이미지 등 각종 정보에 좌표를 비롯한 다양한 지리적 정보를 통합하여 인터넷 지도를 기반으로 다양한 정보를 제공하는 지오타깅(GeoTagging)의 개념으로 확장되고 있다(오충원, 2007)(그림 3).

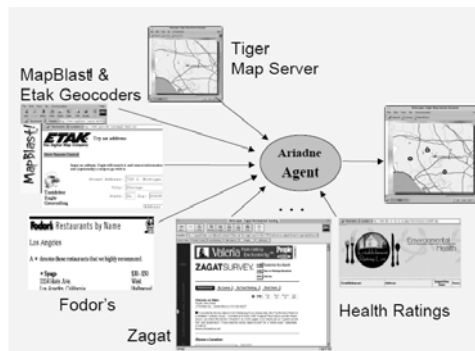
GIS의 응용분야가 많아지면서 NavTech와 TeleAtlas처럼 참조자료와 customized geocoding tool을 제공하는 기업들도 증가하고 있다. 여기에 추가적으로 Google Maps, Map Quest, Yahoo Maps와 같은 웹 기반 지오코딩 엔진들이 많이 사용되고 있으며, 보다 정확한 지오코딩을 요구받고 있다(Paul,

2008).

지오코딩은 주소를 위치정보를 가지는 GIS 데이터베이스로 변환하는 공간분석의 사전 준비에 해당하며, 지리-데이터베이스를 지도화하는 연결고리를 제공하기 때문에 다양한 GIS응용분야에서 지리적 위치를 주소와 매치하기 위한 기능을 필요로 한다(이지영, 김학열, 2006).



〈그림 2〉 주소 지오코딩 개념



〈그림 3〉 web 2.0 기반 지오코딩 개념

출처: Thakkart 등, 2002, Dynamic composing web services from on-line source,

2. 지오코딩 프로세스

지오코딩의 기본적인 방법론은 'Probabilistic record linkage'이다. Probabilistic record linkage는 불확실성 하에서 두 데이터의 정합과정을 의미한다. 이 과정은 데이터가 개별적인지 그룹인지를 판단하여 공통적인 데이터를 표현하고 확인하며, 각 데이터들이 잘 정합되는지 혹은 그렇지 않은지에 대해 점수화하는 fuzzy(퍼지) 논리의 형식을 사용한다. 이 개념은 'Deterministic record linkage' 와 대별되는 것으로, Deterministic record linkage는 두 데이터를 연결할 때 오차가 없다고 가정하며, 두 데이터를 연결시킨다.

이러한 확률적 체계에서 각 속성값은 확률로 측정된 오차에 종속되며 확률적 할당은 인간의 의사결정과정과 흡사하다. Matthew Jaro는 확률적 데이터 연결 방법은 컴퓨터가 연속적이고 효과적으로 공통의 데이터를 찾아내고 평가하는 과정을 되풀이하기 위해서는 필수적이라고 하였다(Nicoara, 2005).

지오코딩은 Probabilistic record linkage 방법론에 기초를 두고 아래의 세 단계를 통해 이루어진다.

1) parsing step

parsing step는 입력된 데이터베이스의 주소를 구조화되도록 변환하고 matching step을 위해 형식을 표준화하는 단계로 입력된 데이터베이스의 주소를 공백과 같은 구분자에 따라 파싱(parsing)한 후, 문자에 대한 규격화가 이루어진다.

이 단계는 다양한 문자와 문자기반의 알고리즘으로 구성되어 있다. 지오코딩을 위한 주소데이터의 주요한 문제는 도로명이 간략화 되어 있거나 (예: street

를 st. st로 작성한 경우) 도로명에 철자가 잘못되었거나(main을 man, marin 으로 작성한 경우), 방향이 복잡하게 조합되어 있는 등 애매한 정보가 포함되어 있는 경우이다.

2) matching step

matching step은 주어진 입력데이터를 참조주소데이터베이스로부터 가장 유사한 참조정보를 찾는 단계이다. matching key(정합 키)에 맞춰 주소의 분할된 요소를 할당한다. 정합기법은 참조 자료에 의존하여 여러 가지 방법으로 실행된다.

3) locating step

마지막으로 locating step은 입력된 주소에 실제 지리적 좌표(x, y)를 결정하는 단계이다. address locator에 입력된 참조데이터와 주소데이터 간의 인덱스 값을 계산하고 각 주소요소와 인덱스 요소가 얼마나 근접하느냐에 따라 매칭점수의 값이 정해진다. 계산된 값을 통해 위치의 잠재적인 후보지를 탐색하여 점수로 환산하여 잠재 후보지를 필터링 한 후 최적의 위치를 결정하며 해당 feature에 좌표를 할당한다.

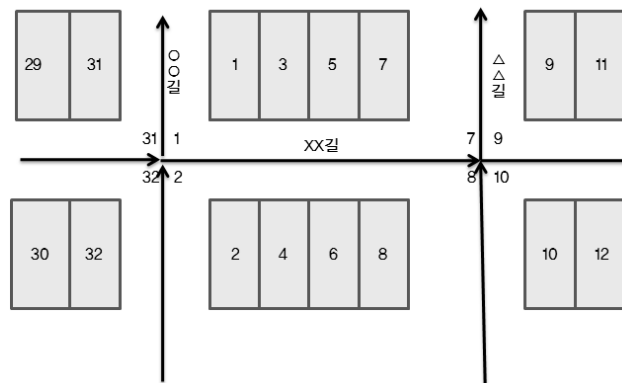
3. 주소모델

3.1 도로망모델

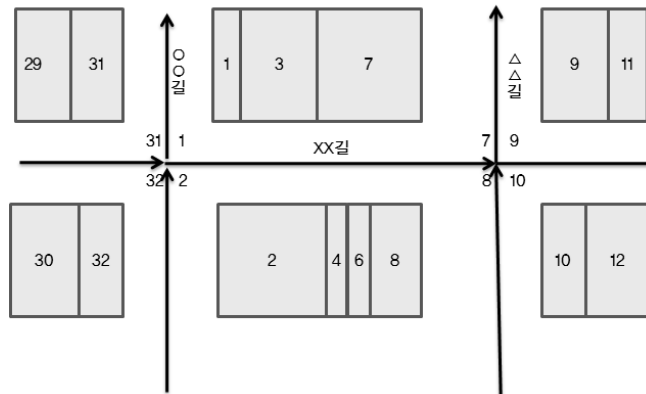
도로망모델은 지오코딩에서 가장 많이 사용되고 있는 주소데이터모델의 도로

망데이터에 기초하고 있다. 이 접근방식에서 도로망은 도로중심선의 세그먼트들로 나타나며, 도로의 양 측면에는 블록과 지번, 도로명이 기재되어 있다. 도로망모델의 지오코딩은 도로명을 매치시킨 후 지번을 포함하는 세그먼트와 선형적 내삽에 기초하여 해당 주소를 좌표상에 point(feature)로 위치시킨다. 이러한 접근방법은 ‘도로 지오코딩’이라 하며, 지오코딩의 주소모델 중 가장 많이 사용되고 있다. 지오코딩을 제공하는 기업 및 GIS S/W에서 대부분 제공하고 있다.

〈그림 4〉는 각 필지들이 동질하게 분포하고 있다고 가정하는 도로망모델이다. 이와 같은 주소모델의 경우 지오코딩의 품질이 높게 나온다. 그러나 현실에서는 〈그림 5〉와 같이 이질적으로 분포하므로 지오코딩의 품질을 낮추는 원인으로 작용하고 있다.



〈그림 4〉 도로망모델 : 동질적



〈그림 5〉 도로망모델 : 이질적

3.2 필지모델

필지는 공간적으로 가장 정확한 주소모델 데이터이다. 필지에 대비한 지오코딩은 필지가 규칙적으로 설정되지 않은 지역에서 유용하게 이용되고 있다.

필지와 도로망모델 간 지오코딩의 차이점은 하나의 필지는 하나의 지번을 가지지만, 하나의 도로 세그먼트(street segment)는 여러 주소의 범위를 가진다는 것이다. 지번을 통한 완벽한 정합은 필지기반 지오코딩에서만 얻을 수 있다는 것을 암시한다. 필지기반 지오코딩은 일반적으로 높은 공간정확도를 가지며, 미국 내 필지와 많은 도시에서 필지기반의 데이터베이스 개발이 이루어지고 있다(Rushton 등, 2006).

〈그림 6〉과 같이 필지모델 지오코딩을 위해서 하나의 필지는 하나의 주소만이 가지므로 모든 주소에 대한 지적DB가 필요하다. 지번주소체계에서 필지기반 지오코딩은 필지별 지번과 연속지적도상의 지번코드를 대조하여 지번자료를 추출한 후 DBF파일로 변환하여 연속지적도와 결합시켜 주소를 매칭한다.

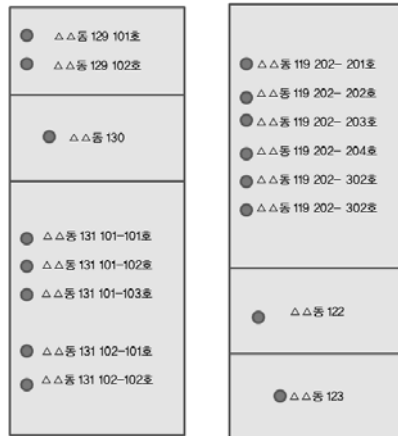
○○동 129	○○동 119
○○동 134	○○동 140
○○동 131	○○동 142
○○동 130	○○동 500-1
○○동 200	○○동 500-2
○○동 204	○○동 504
○○동 205	○○동 507
○○동 208	○○동 510

〈그림 6〉 필지기반 모델

그러므로 지번DB 없이는 지오코딩이 불가능하지만, 지번DB는 자료의 접근성이 낮아 지오코딩을 위한 자료구축에 어려움이 있다.

3.3 address-point 모델

필지기반의 지오코딩의 한계점을 극복하고자, address-point 모델이 등장하였다. address-point 모델은 모든 주소의 마스터주소파일(MAF, master address file)에서 얻어지며, 응급 상황에 대비하고자하는 목적에서 구성된 E911주소 목록의 형식이다. 오스트레일리아와 영국은 국가적 차원에서 address-point 데이터베이스를 개발하였으며, 영국은 영국 국립지리원(Ordnance Survey)에서 개발하여 address-point dataset으로 언급하고 있다. 〈그림 7〉과 같이 address-point는 하나의 필지에 대해서도 여러 개의 주소를 가질 수 있으며, 필지중심점은 필지 내 아무위치에나 위치할 수 있다. 이 주소모델은 개별 아파트 호수, 오피스텔 호수 등과 같은 상세 주소도 지원한다.



〈그림 7〉 address-point 모델

4. 품질평가

품질평가는 지오코딩된 산출물에 대해 참조데이터와의 비교를 통해 정확도를 평가하는 것이다. 많은 연구들은 정합률(matching rate)을 언급하였으나 Whitsel 등(2004)은 정합률의 강조가 잘못된 위치나 잠재적인 오해(misleading)를 불러일으킬 있음을 지적하고 있다. 지오코딩의 전반적인 품질은 완전성, 위치 정확성으로 평가된다.

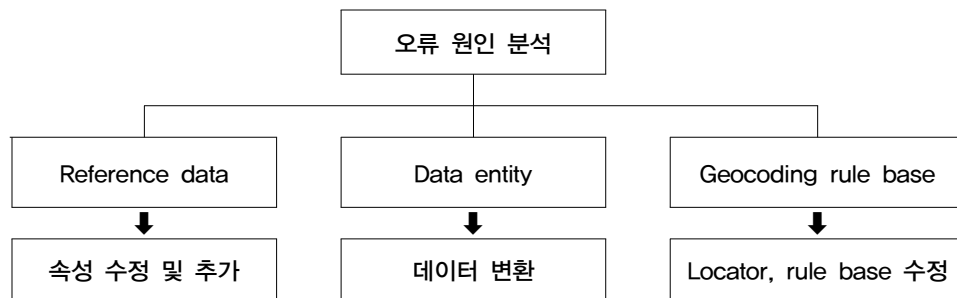
〈그림 8〉은 지오코딩 시 나타나는 오차의 원인을 분석하고 이에 대한 개선방법을 나타내고 있다. 오차의 원인은 크게 reference data 오류와 data entity 오류, geocoding rule base 오류가 있다. reference data 오류는 참조데이터 구축 시 입력될 주소데이터의 속성이 누락된 경우이며, 해당 참조데이터에 속성 데이터를 수정하거나 추가함으로써 품질을 개선할 수 있다. 두 번째 data entity 오류는 입력되는 데이터의 오류로써 주소의 형식을 변환하는 등의 전체

리가 잘못된 경우 이에 해당한다. 이를 개선하기 위해서는 데이터를 특정 형식에 맞추어 변환하는 과정이 필요하다. 세 번째 geocoding rule base 오류는 지오코딩 시 지정하는 locator 및 지오코딩의 rule base가 주소의 형태와 맞지 않기 때문에 발생하는 오류이다. 이를 위해 해당 주소에 맞는 적절한 locator 및 rule base를 설정해주어야 한다.

4.1 완전성(정합률)

완전성은 데이터가 실제 현상을 얼마나 표현하고 있는지를 말한다(정재준, 1998). 해당 주소가 신뢰할 수 있게 지오코딩된 비율이며, 완전성(completeness)이라고도 언급된다. 정합률 또는 신뢰성 있는 주소들의 비율은 지오코딩 품질을 측정하는 의 가장 간단한 방법이다.

Ratcliffe(2004)는 몬테카를로 시뮬레이션을 통해 통계적으로 신뢰 있는 패턴은 정합률이 85%이상이어야 한다고 하였다. 그러나 정합률은 많은 요소에 의존되어 있기 때문에 다양한 정합률을 가진다.



〈그림 8〉 지오코딩 품질 저하원인 및 개선방법

자료 : Nicoara(2005)를 일부 수정함.

4.2 위치정확도

위치정확도는 실제위치와 자료의 위치가 일치하는가를 나타내는 것이다. 위치정확도를 나타내는 요소로는 편의(bias)와 정밀도(precision)이 있다. 보통 편은 오차의 평균으로 나타내며, 정밀도는 평균제곱근 오차로 나타낸다(정재준, 1998).

지오코딩의 위치정확도는 각 지오코딩된 포인트가 그 주소의 참 위치에 어떻게 가까운지를 가리킨다. 몇몇 연구는 지오코딩의 위치정확도를 정량적으로 평가하였다. 예를 들어 Cayo and Talbot(2003)의 연구에 따르면 도시지역의 10%가 약 96m보다 큰 오차가 발생하였고, 5%는 152m보다 크게 오차가 발생하였다. 반면 농촌지역 주소는 2.5~2.9km의 위치오차를 보였다.

지오코딩된 주소의 위치오차는 공간분석적 방법에 반대로 영향을 준다. 위치오차는 클러스터나 경향으로써 공간적 객체를 탐지를 감소시키거나 파라미터 추정의 오차를 증가시키기도 한다.

제2절 주소체계

1. 주소표시방법

주소의 위치표시방법은 크게 지번주소, 도로주소, 격자좌표 체계로 구분할 수 있다. 지번주소체계는 최소행정구역인 동 · 리에 개별필지마다 토지지번을 부여한다. 지번은 토지의 식별과 위치추측을 위해 사용된다. 지번주소방식은 주소 표시 단위동마다 블록을 설정한 다음, 기초번호를 일정규모로 부여하고 이 기초

번호에 따라서 건물에 주소를 부여하는 방법이다(강영옥, 이동연, 1996). 지번 주소방식은 한국과 일본에서만 사용하며, 필지나 블록과 같이 지적성격이 보다 강하다(이지영, 김학열, 2006).

도로망주소체계는 도시를 통과하는 모든 도로에 일정한 명칭을 부여하고 도로의 왼편과 오른편에 건물번호의 범위를 부여하여 일정한 간격을 설정하고 기점으로부터 각 간격에 순차적인 번호를 부여한다. 우리나라의 새주소체계가 이에 해당하며 미국과 동남아시아, 남아메리카, 아프리카 등 전세계적으로 가장 많이 사용되는 주소체계이다.

격자(grid)좌표 혹은 지오코딩(geocoding)주소체계는 토지 내 중심점의 지리적 좌표로서 토지를 구분하여 특정지도 투영법에 의해 좌표상의 위치를 x , y 좌표로 나타내는 방법이다. 예를 들어, 우리나라는 지적좌표계인 TM(Transverse Mercator)도법의 x , y 좌표값이 주소표시 단위가 된다. 수많은 숫자를 사용하여 구성되기 때문에 전산화 체계 속에서 정보의 전달 및 연계 등에는 매우 유용하지만, 인간이 실제 특정 대상물(건물)을 찾는 주소로 사용하기에는 이미지 형성에 기여하는 물리적 요소(통로, 랜드 마크 등)를 갖추지 않아 부적합하다(임재일, 2001).

[표 1] 주소표시방법의 장·단점

구분	지번	도로명	지오코딩
장점	재개발사업, 토지구획정리사업 등을 통해 지번 정리 가능	<ul style="list-style-type: none"> - 블록을 단위로 주소를 표시하므로 획일적이지 못한 도로망에 적합 - 주소단위등을 주소표시의 기본단위로 사용해 거부감이 적음 	토지의 분필·합병에 따라 번호체계가 혼란될 가능성 적음
단점	<ul style="list-style-type: none"> - 토지분할·합병에 따른 변동이 계속적으로 발생하여 비연속적인 지번문제 발생 - 지번정리사업이 실시된 지역과 실시되지 못한 지역 간 지번연계성 결여 	블록 단위 획정이 어려운 골목길을 따라 주소부여시 주변지역과 주소표시 차이 발생	지오코드는 도로명, 가옥형태, 용도지역 등 도시구조와 무관함

자료: 강영욱, 이동연, 1996, 위치정보제고를 위한 주소표시제도 개선방안.

2. 우리나라 주소체계

2.1 지번주소

2.1.1 지번의 형태

지번(地番, parcel number)은 토지의 특정성을 보장하기 위하여 단위구역마다 붙이는 고유번호이다. 지번은 본번(main number)과 부번(sub number)으로 구성되어 있다(지적법 시행령 제3조 제3항). 본번은 지번설정 지역 단위를 붙이는 1,2,3 등 부번을 붙이지 않는 토지번호이다. 반면에 부번은 1976년 4월 시행된 지적법에서 새로 추가된 호칭으로 종전에는 이를 지번 또는 부호하

고 했다. 부변은 -1, -2, -3 등으로 표시된다. 지변은 본변만으로 구성되는 경우도 있고, 본변에 부변을 붙여서 구성하는 경우가 있지만 부변만으로는 독립 지변이 구성되지 못한다(지종덕, 2005). 지변은 본변과 부변의 결합여부에 따라 단식지변과 복식지변으로 분류된다.

단식지변은 본변으로만 구성된 지변형태이다. 단식지변은 표기가 간단하여 결변이 없는 경우 지변으로서 토지의 필지를 추측할 수 있다는 장점을 가진 반면 신규 등록, 등록전환, 분할 등에 의한 지변의 신설이 빈번하게 이루어지면, 지변에 의하여 위치파악이 어렵다는 단점을 가진다.

반면 복식지변은 본변에 부변을 붙여서 구성되는 지변 형태로 지변지역마다 지변의 부여가 완료된 후에 지변의 추가, 분할과 같은 변경이 있을 때 본변에 [의 -], [의 二] 등의 부변을 붙이는 방법을 사용한다. 표기가 복잡하다는 단점이 있지만, 단지단위법에서는 복식지변이 위치추측에 유리하다.

2.1.2 지변설정기준과 방식

1) 설정단위에 따른 분류

지변 부여의 설정단위는 지변설정 지적도 또는 블록의 지변설정 방식에 따라 구분되며 지역단위법, 도엽단위법, 단지단위법으로 구분된다. 지역단위법은 지적도 또는 블록 등의 배열과는 관계없이 지변 설정 지역을 통해 순차적으로 지변을 선정하는 방식으로 우리나라에서는 임야도처럼 축척이 작아 1개의 지변 설정은 1~2장 정도의 도엽에 포용할 수 있는 경우 사용되고 있다.

도엽단위법은 지변설정 지역 내의 지적도 단위도엽의 지변설정이 끝나면 그

다음 도엽으로 옮겨가는 방식으로 지적도 1호부터 도엽의 필지를 연속설정하고 다음 원도에는 먼저 설정된 다음 번호부터 설정하여 먼저 도엽의 지번과 이를 연결시키는 것이다. 우리나라에서 가장 많이 사용되고 있으며 해당 지번 설정 지역의 면적이 비교적 넓고 또한 지적도의 수가 많을 때 주로 사용한다.

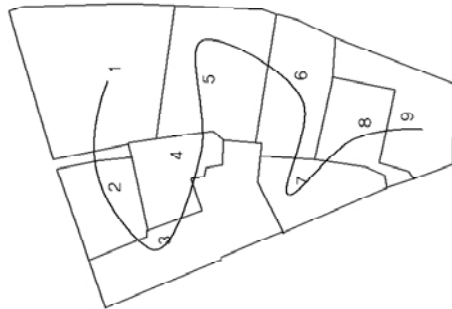
단지단위법은 지적도 배열에 관계없이 여러개의 토지가 1개의 집단을 형성하고 있는 단지마다 지번설정이 끝나면 다른 단지로 옮겨가는 방식이다. 본번 1번의 단지와 본번 2,3,4 등의 단지로 구별한 후 각 단지 내에는 본번을 같이하고 부번만 달리한 복식지번을 붙이기 때문에 지번으로 토지위치를 찾아내는데 편리하다. 우리나라에서 구획된 시가지나 경지정리지구등에 채택하고 있다.

2) 진행방식에 따른 분류

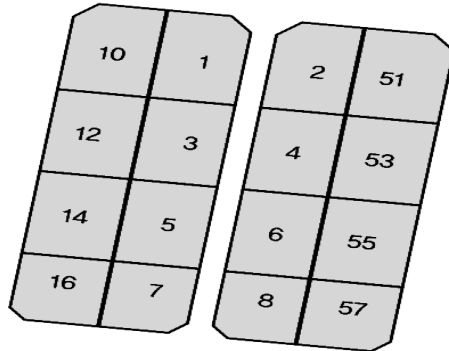
지번은 구역 획정 후 각 구역별로 설정해야한다. 구역별 지번 설정방식은 사행식, 기우식, 절충식이 있다. 사행식(蛇行式)은 지번을 붙이는데 가장 널리 쓰이는 방식으로 뱀이 기어가는 것처럼 토지의 진행순서를 따라가는 형상의 설정 방법이다. 이것은 토지 배열이 불규칙한 때에 편리한 방식으로 연속적으로 설정했을 경우 인접 지번의 일렬로는 연결되었지만, 인접 지번번호가 넘어가는 결점이 있다. 우리나라에서는 대부분이 사행식으로 주소가 설정되어 있다.

기우식(基偶式)은 도로 등을 중심으로 하여 왼편은 홀수(기수, 基數)로 하고 오른편은 짝수(우수, 偶數)로 지번을 설정하는 방식으로 시가지 등에 이용할 때 토지소재를 추측할 수 있는 이점이 있으며 교호식이라고도 한다. 우리나라에서는 부분적으로 구획정리 등이 완료된 지역에서 기우식을 사용하고 있다.

절충식은 토지의 배열이 고르지 못한 지역에서 선택하는 방식으로 사행식, 기우식 등을 적당히 선택해 지번을 설정하는 방법이다. 우리나라에서 절충식은 지적도 1도엽에 포용하는 토지가 불규칙하게 배열된 것과 비교적 정연하기 배열된 것이 혼재하는 지역에 대해 불규칙한 지역은 사행식으로 지번을 설정하고, 정연한 구역은 기우식 혹은 단지식으로 설정하여 사용하고 있다.



〈그림 9〉 지번 설정방식 : 사행식



〈그림 10〉 지번 설정방식 : 기우식

2.1.3 지번체계의 문제점

우리나라의 주소표시방법은 지번에 기초하기 때문에 지번의 체계가 무질서하면 주소표시도 문제점을 갖게 된다. 지번의 무질서를 배제하더라도 일반인이 주소를 찾을 때는 토지보다는 건물이나 주요도로를 중심으로 주소를 인지하게 되므로 지번은 체계가 도시구조를 반영하지 못하면 주소표시의 단위로서 문제점을 갖게 된다.

2.2 도로명주소(새주소)체계

2.2.1 새주소 사업 도입배경 및 기대효과

도로명주소(새주소)는 OECD 등 주요 선진국들과 같은 도로방식에 의한 주소체계로서, 도로마다 도로명을 부여하고 도로의 진행방향에 따라 왼쪽 건축물에는 홀수번호를 오른쪽 건축물에는 짝수번호를 순차적으로 부여하여 사용하는 주소를 말한다(행정자치부, 2007). 새주소 표기는 지번 대신 도로명, 건물번호 및 상세주소로 표기하며 이에 대한 예는 [표 2]와 같다.

지번주소체계는 건물의 소재지와 토지의 소재지를 동일하게 표시하는 것으로 산업화·도시화의 진전으로 도시의 구조가 복잡해지고 거대화됨에 따라 주소의 기능 중 위치 표시의 기능의 여러 가지 문제점이 지적되었다. 대표적인 예로는 <그림 11>과 같이 인접한 지역임에도 불구하고 번지가 완전히 다른 경우가 있으며, <그림 12>와 같이 한 건물이 여러 지번에 걸쳐있는 경우도 있다.

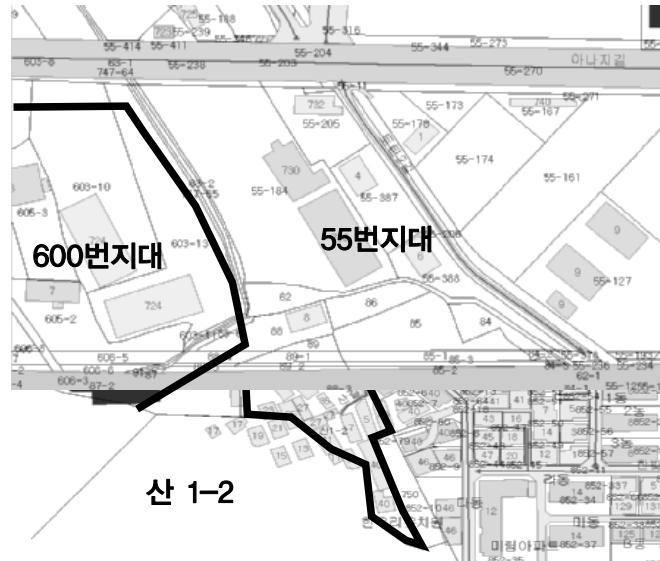
[표 2] 새주소 표기방법(예)

구 분	지 번 주 소	새주소(도로명주소)
단독주택	서울특별시 은평구 갈현동 510-18	서울특별시 은평구 메밀밭길 10 (갈현동)
업 무 용 빌 딩	서울특별시 종로구 수송동 146-1 이마빌딩	서울특별시 종로구 중학천길 42 (수송동)
공동주택	서울특별시 노원구 상계동 624 주공아파트 ○○동 ○○호	서울특별시 노원구 늘푸른 1길 14 ○○동 ○○호(상계동, 주공아파트)

출처: 행정자치부, 2007, 묻지않고 찾아가는, 자신있는 발걸음.

〈그림 12〉에서 진한 선 안의 음영부분 건물들은 여러 개의 지번을 가짐으로써 대표지번을 선정함에 있어 혼선이 있었다. 이러한 문제점을 해소하기 위하여 도로방식에 의한 주소표시방법으로 변경하였다.

새주소체계는 건물에 번호를 부여하기 위해 골목길까지 일정한 원칙에 의거하여 도로명과 건물번호를 부여한다. 이 원칙을 이용자가 숙지하면 보다 쉽게 목적지를 찾을 수 있다. 그리고 지번에 의한 주소제도보다 상대적 위치파악이 가능하다.



〈그림 11〉 지번주소의 문제점1 : 지번배열 순서가 불합리하여 순차적 위치찾기가 어려운 경우
 (계양구 서운동 55번지대 인근 600번지대 혼재)
 출처: 행정자치부, 2007, 묻지않고 찾아가는, 자신있는 발걸음.



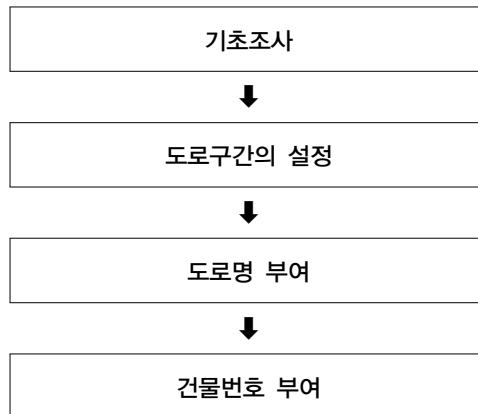
〈그림 12〉 지번주소의 문제점2 : 한 건물이 여러 지번에 걸쳐있는 경우
 (인천광역시 계양구 효성동 산 59-5 외 6필지)
 출처: 행정자치부, 2007, 묻지않고 찾아가는, 자신있는 발걸음.

주소체계 변화로 약 9,180종의 주민등록, 등기부 등 각종 공부 상의 주소가 변경되고, 재난·방범·의료 등 공공부문은 물론 자동차 네비게이션, 모바일 등 민간 IT분야의 서비스개선에 기여함으로써 유비쿼터스시대 최적의 위치정보로 활용될 것으로 전망된다(행정자치부, 2007).

2007년 4월 「도로명주소 등 표기에 관한 법률」이 발효되었으며, 전국적으로 사업이 완료되는 2009년 말부터 본격적으로 사용가능하며, 2011년 12월 31일까지는 현재의 지번주소와 새로운 도로명주소가 병행 사용된다.

2.2.3 도로명주소의 형태

도로명주소는 도로명과 건물번호로 구성되어 있다. 도로명주소의 부여 순서는 <그림 13>과 같다.



<그림 13> 도로명 주소 부여 순서

자료: 도로명주소 등 표기에 관한 법률 제8조 제1항.

1) 도로구간 및 도로명

도로구간은 도로 폭에 따라 대로, 노(路), 길로 구분된다. 대로는 도로의 폭이 40미터 또는 왕복 8차로 이상인 도로이며, 노(路)는 도로의 폭이 12미터 이상 40미터 미만이거나 왕복 2차로 이상 8차로 미만인 도로이다. 대로와 노(路) 외의 도로는 길로 구분된다. 도로구간은 도로의 연속성을 유지하면서 최대한 길게 설정하고, 다른 도로구간과 중복되지 않으며, 기점에서 종점까지 연결되도록 설정하며 가급적 직선에 가깝도록 한다. 일시적인 도로에는 도로구간이 설정되지 않는다.

도로구간이 설정되면 각 도로구간에 대해 도로명을 부여한다. 도로명은 10일 이상의 기간을 정하여 공보와 주민의 의견을 수렴한 후 부여된다. 도로명은 각 지역에 따라 지역의 특성, 역사성, 위치예측성, 영속성, 지역주민의 의견 및 지명을 종합적으로 고려하여 결정된다.

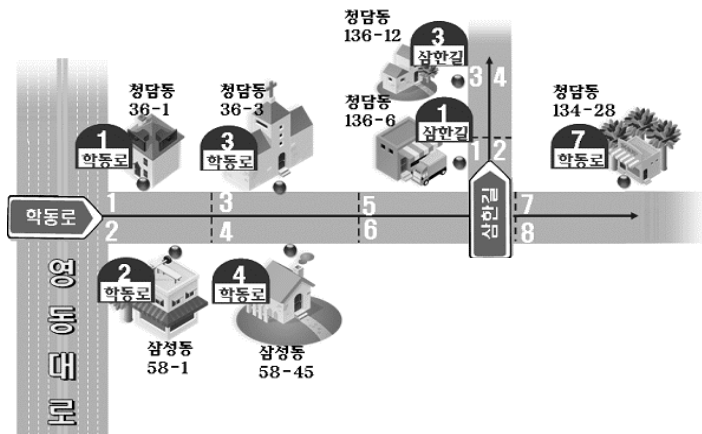
2) 건물번호

건물번호는 건축물대장에 등재된 주된 건물에 부여된 번호를 말한다. 지번주소가 토지에 부여했던 것과 달리, 도로명주소의 건물번호는 건축물에만 번호가 부여되며 기초번호와 가지번호로 구성된다.

건물번호는 해당 건축물들의 주된 출입구에 접하고 있는 도로구간에 대해 기초번호를 부여하도록 되어 있으며, 출입구가 여러 개 있는 경우에는 해당 건축물 등의 중앙에 가까운 기초번호나 교차로에 위치한 출입구에 접하고 있는 도로구간의 기초번호를 기준으로 부여하도록 되어 있다.

기초번호는 지번주소에서 본번에 해당하는 요소로, 도로구간에 대해 일정간격

을 나눠서 부여한다. 기초번호는 도로변에 위치한 건축물과는 관계없이 20미터 간격으로 결정하며, ‘길’에 해당하는 도로는 10미터 간격이다. 모든 도로의 기초번호는 도로의 기점에서 종점방향으로 왼쪽은 홀수번호가 부여되며 오른쪽은 짝수번호를 부여한다. 하나의 건물이 기초번호에 있으면 기초번호가 부여되며, 여러 개의 건물이 하나의 기초번호에 있을 경우 가지번호가 부여된다. 한 건물의 출입구가 여러 개이면 해당 기초번호의 중간 값 또는 첫 번째 기초번호가 건물번호로 부여된다. 만약, 건물이 둘 이상의 도로변에 출입구가 접해 있을 경우, 도로폭이나 길이가 더 긴 도로에 기초하여 기초번호가 부여되며, 도로구간이 대등한 경우(예: 대로-대로 등) 교통량이 많은 도로변에 접한 출입구의 기초번호에 건물번호가 부여된다. 하나의 건물번호를 부여한 공동주택 등이 단지도로가 아닌 통과도로에 의해 여러 구역으로 나누어진 경우에는 구역별로 주된 출입구가 접한 도로의 기초번호를 건물번호로 부여 할 수 있다.



〈그림 14〉 새주소체계의 건물부여 방식
 자료: 새주소안내시스템(<http://www.juso.go.kr>)

가지번호는 지번주소에서 부번에 해당하는 요소로 ‘-’ 표시는 ‘의’ 로 읽고, 가지번호 뒤에 ‘번’ 을 붙여 읽는다. 가지번호가 부여되는 도로구간인 종속구간은 10미터 또는 5미터 간격으로 설정하며 하나의 기초번호 범위 안에 둘 이상의 건축물이 있을 경우, 해당 도로구간의 기점에서 종점방향으로 두 번째 건물부터 가지번호를 붙여 건물번호가 부여된다. 종속구간의 기초번호는 주된 도로구간과 연결된 지점의 가지번호를 붙이며, 가지번호는 기점에서 종점방향으로 차례대로 번호가 부여된다.

제3장 지오코딩 적용 실험

제1절 실험 계획 및 연구지역 설정

이 장에서는 위의 이론적 배경을 바탕으로 새주소체계에 따라 지오코딩의 가능성을 확인하고, 품질평가를 통해 도로중심선 기반 지오코딩의 품질개선 방안을 모색해보고자 한다.

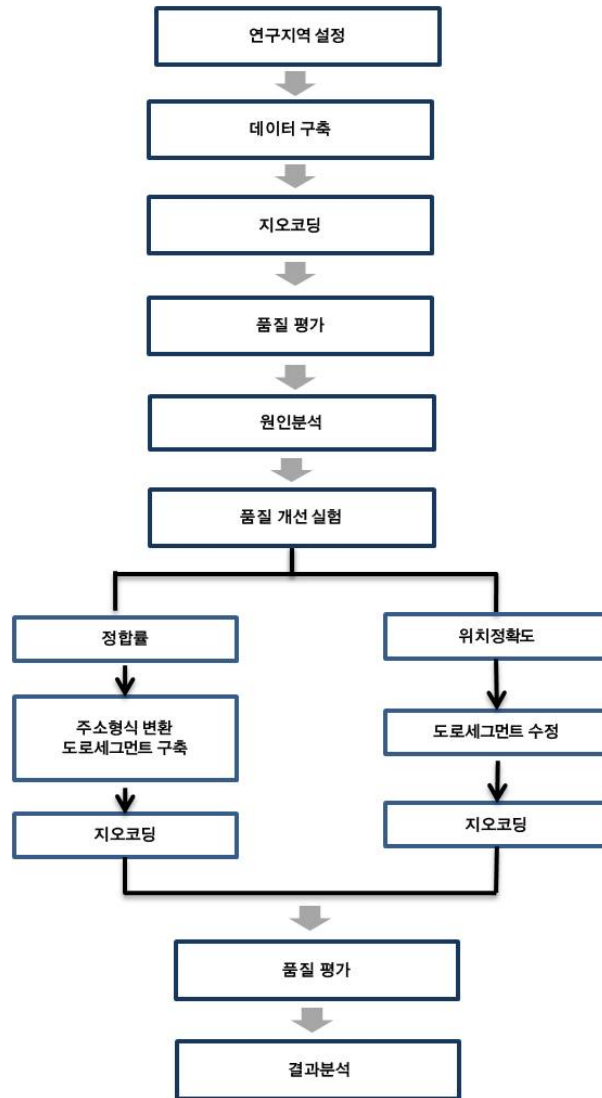
지오코딩 실험을 위해 서울특별시 성북구의 동선동일대를 실험지역으로 선정하였다. 동선동은 동선동 1,2,3,4,5가로 구분되어 있으며, 본 연구에서는 1,2가 전체와 3가의 일부를 선정하였다(그림 15).



〈그림 15〉 실험지역 행정구역

실험지역으로 지정한 동선동일대는 동선동 1가의 하나로 거리를 중심으로 한 상업지역과 이 밖에 거주지역이 혼재되어 있으며, 거주지가 잘 정리된 지역과 그렇지 않은 지역 역시 혼재되어 있다. 또한 동선동 일대는 혼합형 도로망으로써 격자형과 자연발생형이 혼합되어 있는데, 동선동 2,3가는 격자형 도로망이고, 동선동 1가는 격자형과 자연발생형 도로망이 혼합되어 있다.

지오코딩을 위해 새주소체계의 주소데이터 및 도로 세그먼트(street segment)를 구축하여 지오코딩하였다. 지오코딩이 된 데이터는 품질평가를 실시하여 결과를 분석하고, 이러한 분석을 기초로 품질개선을 실험을 실시하였다. 실험은 정합률 개선 실험과 위치정확도 개선 실험으로 구분하였다. 실험 후 지오코딩의 결과에 대해 품질평가를 실시하고, 마지막으로 실험 전 후의 품질평가를 실시하는 결과분석을 하였다. 전체적인 실험흐름은 <그림 16>과 같다.



〈그림 16〉 실험흐름도

제2절 데이터 구축 및 지오코딩

1. 도로 세그먼트

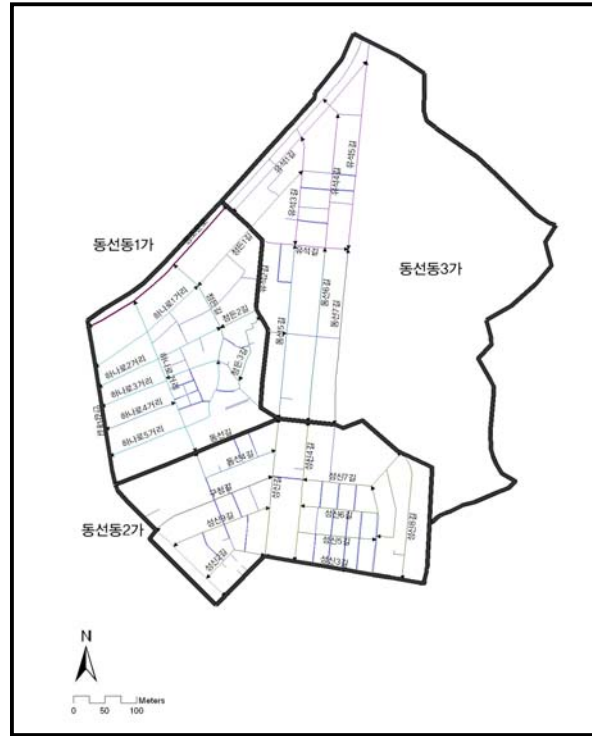
도로 세그먼트는 도로 지오코딩을 위한 기초데이터이다. 도로 세그먼트는 다음과 같은 순서로 구축하였다. 도로 및 지역을 선정한 후 도로중심선을 구축하여, 구축된 도로 세그먼트에 도로명 및 도로범위를 입력하였다.

- ① 구축 대상도로 설정
- ② 도로중심선 구축
- ③ 도로명 및 도로범위 입력

본 연구에서는 서울특별시 성북구 동선동 일대를 대상 지역으로 선정하였으며, 대상 도로는 동선동 1,2,3가 일대로 구청길 등 33개이다. 도로중심선은 1:1000 수치지형도에서 추출하였으며, 추출된 도로중심선에 도로명과 도로범위를 입력하였다. 도로 세그먼트에 입력될 정보는 왼편과 오른편에 대한 건물번호의 시작과 끝의 범위로 새주소안내시스템에서 취득하여 [표 3]과 같이 도로 세그먼트를 구축하였다. 도로 세그먼트 내 정보는 ArcMap에서 도로망 셰이프파일(shape file)과 조인하였다(그림 17).

[표 3] 기초번호의 도로 세그먼트

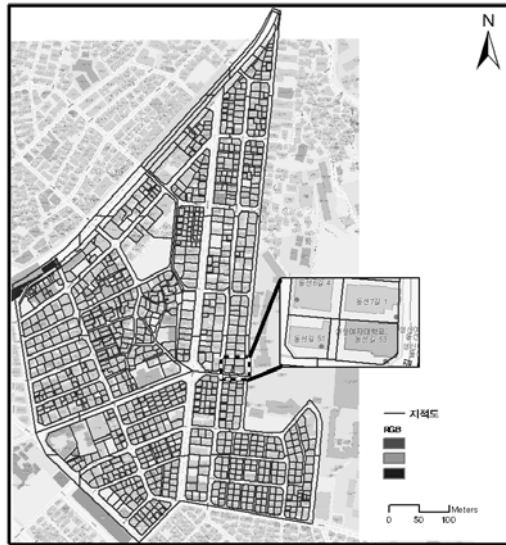
도로명	Left_from	Left_to	Right_from	Right_to
하나로거리	1	41	2	42
하나로5거리	1	37	2	38
하나로4거리	1	13	2	13
하나로3거리	1	25	2	26
하나로2거리	1	37	2	38
하나로1거리	1	41	2	42
정든길	5	33	6	34
정든3길	1	23	2	24
정든2길	1	21	2	22
정든1길	1	25	2	26
유석길	1	25	2	26
유석5길	1	59	0	0
유석4길	1	41	2	42
유석3길	1	35	2	36
유석2길	1	25	2	26
유석1길	1	49	2	50
안감내길	0	0	152	204
성신길	23	65	24	66
성신9길	1	35	2	36
성신8길	0	0	19	1
성신7길	1	21	2	22
성신6길	1	27	2	28
성신5길	1	41	2	42
성신4길	1	43	2	44
성신3길	1	31	2	32
성신2길	1	15	2	16
동소문로	0	0	170	188
동선길	17	53	18	54
동선7길	1	31	0	0
동선6길	1	33	2	34
동선5길	1	41	2	42
동선4길	1	27	2	28
구청길	17	39	18	40



〈그림 17〉 실험지역 도로 세그먼트

2. 참조자료

참조자료는 지오코딩을 위한 정확도 평가를 위해 필요하다. 참조자료로서 성북구 동선동 일대에 대한 지적도와 새주소안내지도가 사용되었다. 지적도에는 각 동에 대한 지번정보가 저장되어 있다. 새주소안내지도는 새주소안내시스템의 '새주소 안내'에서 동선동일대를 캡처한 그림파일에 좌표정보를 부여한 파일이다. 새주소안내지도에는 각 건물의 주출입구가 표시되어 있다. 지적도와 새주소안내 지도파일을 중첩하여 참조자료로 구축하였다(그림 18).



〈그림 18〉 참조자료 : 새주소안내지도 및 지적도

3. 주소데이터 수집 및 변환

실험을 위한 주소데이터는 도로명주소이며, 새주소안내시스템에서 제공하는 ‘14차 테스트용 새주소와 지번 매칭 테이블’을 이용하였다. 새주소안내시스템에서 제공하는 주소는 MDB(Access 2000 통합버전)형태로 약 한 달을 주기로 테스트용 새주소와 지번의 매칭테이블에 대해 전국³⁾의 데이터를 제공하고 있다. 지번주소와 도로명주소를 제공하며, 법정코드(10자리), 시·도, 시·군·구, 읍·면·동, 리, 산 여부, 지번 본번, 지번 부번(이하 지번주소 정보), 도로명, 새주소 본번, 새주소 부번, 건물명, 상세건물명(이하 도로명주소 정보)의 속성을 제공한다.

3) 2008년 12월 현재, 새주소사업이 44%(지역기준, 103개 완료, 129개 추진 중)되었기 때문에, 완료된 지역에 한하여 제공된다.

이렇게 제공된 주소데이터는 지오코딩 소프트웨어에서 제공하는 주소 형식으로 변환해야 지오코딩이 가능하다. 현재 우리나라에서 사용하고 있는 주소표시 형식은 아래와 같다.

시 · 도 / 구 · 군 / 동 · 리 / 지번
 시 · 도 / 구 · 군 / 도로명 / 건물번호

그러나 ArcGIS Street map에서는 다음과 같이 주소를 인식하므로 아래와 같이 변환해야한다.

지번 / 동 · 리 / 구 · 군 / 시 · 도
 건물명 / 도로명 / 구 · 군 / 시 · 도

변환한 주소의 형식은 [표 4]와 같으며, -(하이픈)은 ArcGIS는 공백이 구분자의 역할을 하기 때문에 사용하지 않았다.

2008년 11월 기준 연구지역 내에서 새주소가 부여된 주소는 총 912개로, 589개는 기초번호만이 존재하는 주소이고, 323개는 기초번호와 가지번호가 부여된 주소이다.

[표 4] 주소표시 형식 변환(예)

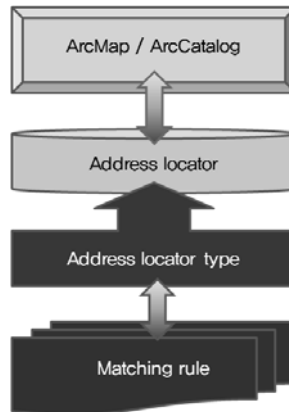
구 분	기존(현행)	ArcGIS 인식
기초번호	동선길 5	5 동선길
기초번호, 가지번호	동선길 54-7	54 7 동선길

4. 지오코딩

4.1 address locator 설정

address locator(주소 탐지기)는 위치에 대한 특수한 속성과 파라미터에 대한 정보를 입력받아 입력주소의 형태를 해석함으로써 주소를 좌표로 변환해주는 역할을 한다(그림 19).

ArcGIS에서는 ArcCatalog와 ArcMap에서 address locator를 만들 수 있으며, 주소의 형태에 따라 다양한 address locator를 생성하여 지오코딩 할 수 있다. ArcGIS에서는 11가지의 address locator를 제공한다(표 5). 본 연구에서는 새주소체계에서 기초번호 부여 방식이 도로를 중심으로 한 기우식으로 건물 번호를 부여하였기 때문에 US Streets를 선택하였다.



〈그림 19〉 address locator의 역할

자료: ESRI(<http://www.esri.com>)

[표 5] ArcGIS Street map에서 제공하는 address locator type

type	geometry	데이터 표현	주소 검색 매개변수	예	응용
US Streets	선	도로의 양쪽에 대한 주소 범위의 세그먼트	단일 필드에 있는 모든 주소요소	320 Madison st.	도로의 특정 측면의 집 찾기
US Alphanumeric	선	격자 구역의 정보가 있는 주소	단일필드에 있는 모든 주소요소	N2W1700 County Rd.	Illinois, Wisconsin의 일부지역에서 사용
US Hyphenated	선	주소내의 교차로 정보	단일필드에 있는 모든 주소요소	105-30 Union st.	Queens와 New York에서 사용
US One Range	선	각 도로의 세그먼트에 대한 하나의 범위	단일필드에 있는 모든 주소요소	2 summit Rd.	도로의 측면이 요구되지 않는 집 찾기
US one Address	점, 면	주소를 나타내는 각 피쳐	단일필드에 있는 모든 주소요소	71 Cherry Ln.	address-point, 필지, 건물 찾기
Single Field	점, 면	특정장소, 랜드 마크, 관심지점을 나타내는 각 피쳐	사용자가 지정하는 변수 한 개	Cabrillo college	장소명이나 랜드마크 찾기
US Cities with State	점, 면	주 내 도시	주 나 도시명, 약어	Rice, Wa	한 국가 내에서 특정 도시 찾기
World Cities	점, 면	국가 내 도시	도시명과 국가명	Lima, Peru	전세계 특정 도시 찾기
Zip 5 Digit	점, 면	우편번호지역이나 중심점	5자리 우편번호	22066	특정 우편번호 위치 찾기
Zip+4	점, 면	우편번호 4자리 지역이나 중심점	5자리 우편번호와 별도의 필드에 4자리 확장	96822, 2323	특정 우편번호 위치 찾기
Zip+4 Range	점, 면	우편번호와 4 자리 숫자를 나타낸 각 피쳐	5자리 우편번호와 별도의 필드에 4자리 확장	63703, 0078	특정 우편번호 위치 찾기

자료: ESRI(<http://www.esri.com>)

4.2 정합

정합은 지오코딩하기 위해 주소데이터에 좌표를 할당해주는 과정이다.

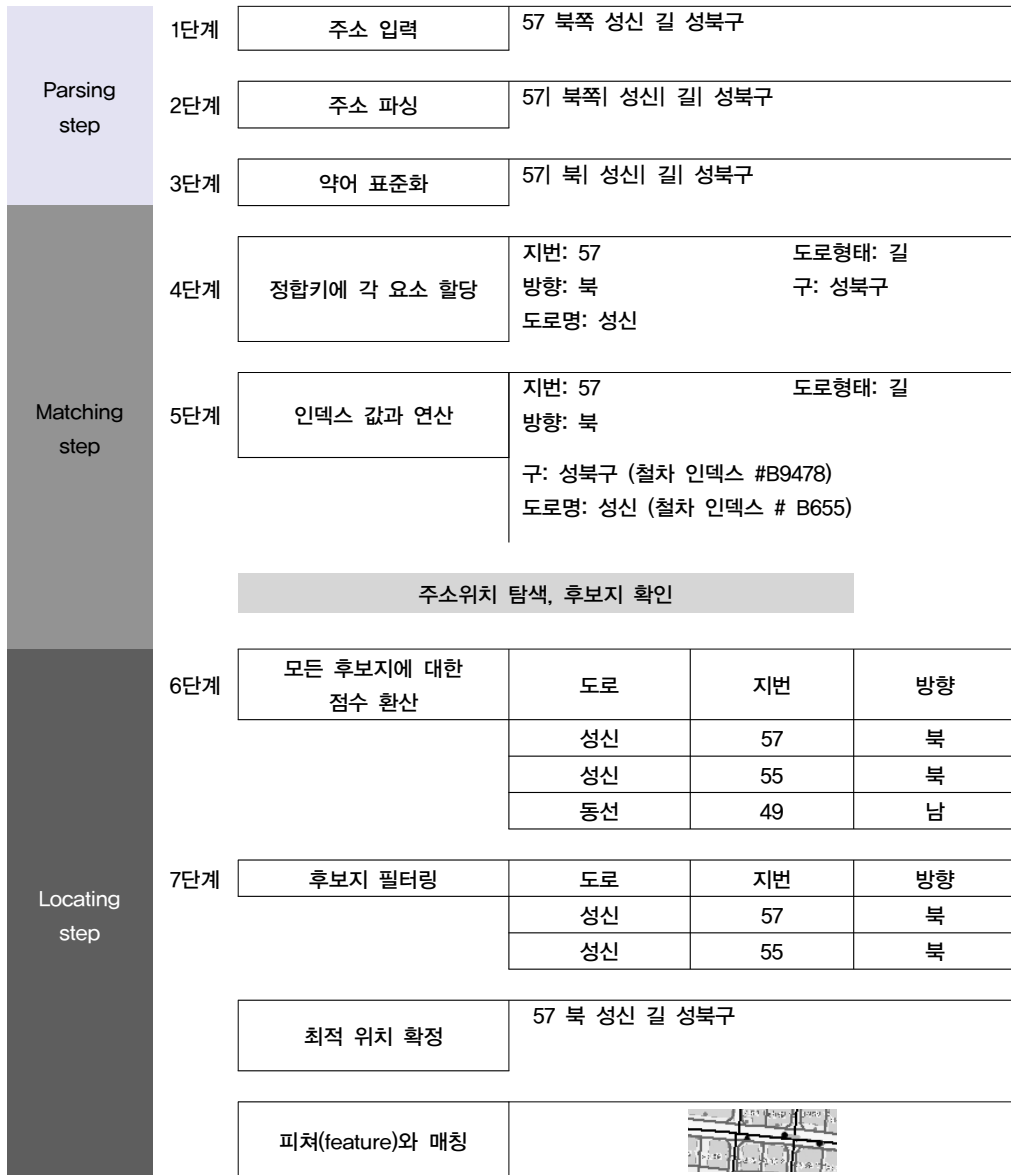
ArcGIS의 Street map을 이용하여 지오코딩을 하였다. ArcGIS의 지오코딩은 크게 세 단계로 나누어 진행된다(그림 20). parsing, matching, locating step을 통해 주소는 좌표를 가진 데이터로 변환된다. ArcGIS는 지오코딩 시 [표 6]과 같은 옵션을 제공한다. 본 연구에서는 각 도로명이 모두 정확하게 매칭되어야 하기 때문에 철자민감도를 100으로 설정하였다. 왼편과 오른편의 건물번호를 구분하기 위해 side offset은 1m로 지정하였으며 end offset은 0으로 지정하였고, 후보의 결과값이 동점일 경우에는 정합하지 않는 옵션을 선택하였다.

[표 6] ArcGIS의 지오코딩 옵션

분류		선택 가능 범위
matching options	철자 민감도 (spelling sensitivity)	0 ~ 100
	최소 후보 점수 (Minimum candidate score)	0 ~ 100
	최소 정합 점수 (Minimum matching score)	0 ~ 100
output options	side offset	—*
	end offset	0 ~ 50
	후보의 결과 값이 동점일 때 정합여부	Yes/ No
output fields	x, y 좌표	Yes/ No
	표준화된 주소	Yes/ No
	참조데이터 ID	Yes/ No
	(도로 세그먼트로부터 위치한) 비율	Yes/ No

* 길이 단위에 대해 사용자가 값을 지정할 수 있음.

자료: ESRI(<http://www.esri.com>)



〈그림 20〉 ArcGIS Street map의 매칭 알고리즘

자료: ESRI(<http://www.esri.com>)

제3절 품질평가

1. 정합률

위의 과정을 통해 지오코딩 된 포인트 데이터에 대해 품질평가로써 정합률과 위치정확도를 평가하였다. 정합률은 전체 912개의 주소에 대해 지오코딩 한 결과 589개의 주소가 매칭되었다(표 7). 매칭된 주소데이터를 분석한 결과 589개는 모두 기초번호로써 가지번호는 매칭되지 않았다.

2. 위치정확도

위와 마찬가지로 주소데이터 중 정합된 주소는 기초번호에 해당하는 것으로, 기초번호에 대해서만 위치정확도를 평가하였다. <그림 21>과 같이 지오코딩을 통해 할당된 포인트와 참조데이터의 주소범위와 거리를 측정함으로써 위치정확도를 평가하였다.

지오코딩된 589개의 주소데이터 중 278개가 해당 주소범위에 할당되어 오차가 없이 매칭되었으나, 311개는 해당 주소범위를 벗어나면서 오차가 발생하였다(표 8, 그림 22).

오차가 발생한 주소데이터의 평균 위치정확도는 6.88m이며 표준편차 6.12, 최소값 0.10, 최대값 40.81이다. 위치정확도 분석 결과 성신9길 부근이 가장 낮았으며, 정든1길과 동선동길 부근 교차로에서 위치정확도도 낮은 것으로 나타났다. 성신9길 부근은 성신여대의 기숙사가 위치하고 있는 지역으로 다른 주소에 비해 넓은 면적을 차지하고 있기 때문에, 다른 주소 포인트들이 밀려나면

서 오차가 크게 발생하였다(그림 23). 또한 위치정확도가 낮게 나온 지역인 동선길 부근 교차로는 도로범위는 교차로 중심부에서부터 시작하지만, 실제 주소는 해당 도로의 시작부분부터 주소가 지정되기 때문에 해당 교차로 부근에서 오차가 크게 나타나는 것으로 분석되었다(그림 24).

[표 7] 전체주소 : 정합률

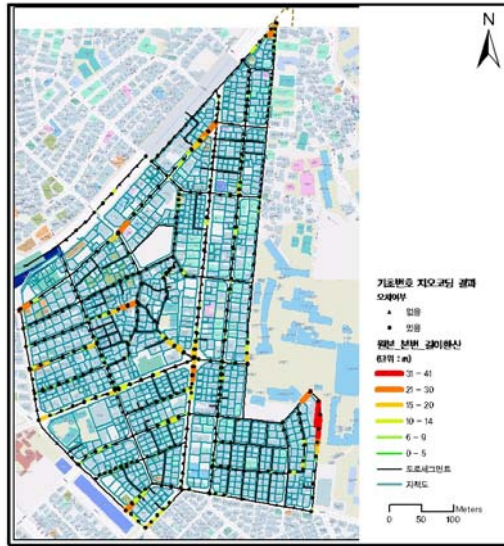
구 분	전체	기초번호	가지번호
주소 개수	912	589	323
매칭	589 (93.1)	589 (100)	0 (0)
매칭 안됨	323 (35.41)	0 (0)	323 (100)

[표 8] 전체주소 : 기초번호의 위치오차 통계량

평균	6.88
표준편차	6.12
중앙값	4.99
최빈값	2.60
최소값	0.10
최대값	40.81
범위	40.71



〈그림 21〉 위치정확도 측정방법



〈그림 22〉 전체 주소 : 위치정확도



〈그림 23〉 전체 주소 : 위치정확도 분석(성신9길 부근)



〈그림 24〉 전체 주소 : 위치정확도 분석(동선길 부근 교차로)

제4장 지오코딩 품질 개선 실험

위의 지오코딩 실험결과를 통해 지오코딩의 품질 개선이 필요한 것으로 나타났다. 지오코딩의 품질 개선 실험은 정합률을 높이기 위한 실험과 위치정확도를 개선하기 위한 실험으로 구분하여 진행하였다.

제1절 정합률 개선 실험

새주소를 지오코딩 할 때, 기초번호를 중심으로 구축한 도로 세그먼트에서 가지번호의 주소가 매칭이 되지 않는 결과가 나타났다. 그러나 기초번호와 마찬가지로 가지번호에 대한 도로 세그먼트를 구축하면 정합률을 개선할 수 있을 것이다. 그러므로 정합률 개선을 위해서는 도로 세그먼트를 구축 및 주소데이터의 변환이 필요하며 구축된 도로 세그먼트에 맞도록 address locator를 재설정하는 작업이 필요하다.

1. 가지번호(전체)

1.1 도로 세그먼트

정합률 개선을 위해 가지번호의 도로 세그먼트 구축이 필요하다. 가지번호의 경우 기초번호와 달리 도로 세그먼트 구축 시 유석1길 27-1과 같이 도로명과 기초번호가 함께 조합되므로 이를 반영할 수 있는 도로 세그먼트의 구축이 필

요하다. 또한 가지번호의 경우 기초번호와 달리 기점에서 종점까지 한 방향으로 연결되어 주소가 지정되므로 시작과 끝의 범위가 필요하다. [표 9]는 위의 사항을 반영한 도로 세그먼트이며, 본 연구를 위해 총 60개의 가지번호 도로 세그먼트를 구축하였다.

[표 9] 가지번호의 도로 세그먼트

도로명	from	to	도로명	from	to
안감내길 172	1	1	성신5길 15	1	4
성신길 54	1	6	정든길 15	1	6
동선길 54	1	7	성신2길 14	1	10
동선길 40	1	4	정든3길 14	1	2
유석4길 40	0	0	유석3길 14	1	4
동선길 38	1	6	유석4길 14	1	2
구청길 36	1	2	하나로1거리 14	0	0
구청길 34	1	8	성신3길 13	1	4
하나로2거리 32	1	3	성신2길 13	1	2
구청길 31	1	3	성신6길 13	1	7
구청길 30	1	4	유석길 12	1	14
정든길 30	1	5	유석1길 12	1	1
성신길 29	1	2	성신3길 11	1	5
성신4길 29	1	1	성신5길 11	1	4
정든길 28	1	4	정든길 10	1	4
성신길 27	1	1	성신6길 9	1	8
유석3길 26	1	3	정든길 9	1	8
성신3길 25	1	8	성신2길 8	1	7
하나로3거리 22	1	12	동선4길 8	1	5
유석4길 22	1	4	유석3길 8	1	5
성신3길 21	1	8	성신5길 7	1	6
정든길 21	1	1	하나로거리 7	1	11
정든1길 21	1	1	하나로거리 7	1	4
유석3길 20	1	5	유석3길 7	1	2
동선5길 20	0	0	성신2길 6	1	7
성신5길 19	1	4	성신8길 5	1	6
유석4길 18	1	4	성신6길 5	1	4
성신3길 17	1	7	성신5길 3	1	3
하나로거리 17	1	4	하나로거리 3	1	6
유석1길 16	1	1	하나로거리 1	1	6

1.2 지오코딩

구축된 도로 세그먼트와 매칭할 수 있도록 주소데이터의 형식을 변화시켜야 한다. 예를 들어, '53 1 동선길'이라는 주소는 '1 동선길53'의 형식으로 변화해야 한다. 즉, 동선길53이 하나의 도로 세그먼트로 구축되었음을 의미한다. 본 실험을 위해 323개의 주소를 위와 같은 형식으로 변환하였다.

위에서 구축된 도로 세그먼트, 변환된 주소, address locator를 이용해 ArcGIS Street map으로 매칭하였다. 위의 기초번호와 마찬가지로 철자민감도는 100으로 설정하였다.

1.3 품질 평가

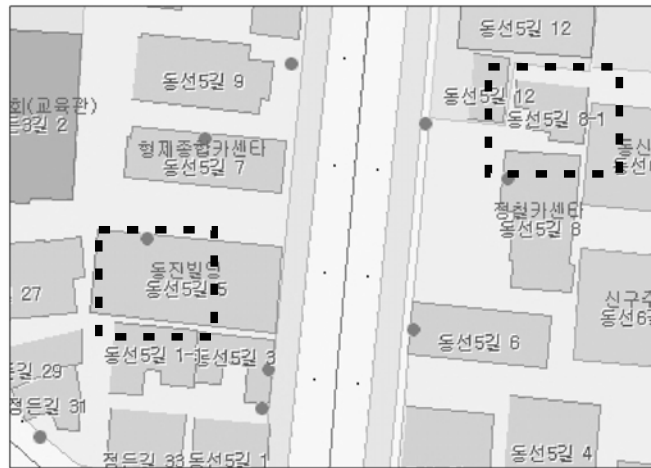
가지번호에 대해 지오코딩을 실시한 결과 323개의 가지번호 중 266개의 주소가 매칭되었고, 57개의 주소는 매칭되지 않았다(표 10).

[표 10] 가지번호 : 정합률

구 분	가지번호
주소 개수	323
매칭	266 (82.35)
매칭 안됨	57 (17.65)

57개의 매칭되지 않은 주소들을 분석한 결과, 가지번호 도로 세그먼트 상에 없는 주소들이었다. 이 주소들은 가지번호 도로 세그먼트가 아닌 기초번호 도로 세그먼트 상에 존재하기 때문이다. 이러한 가지번호가 발생하는 원인은 새주소가 부여될 때 가지번호가 부여되는 원칙 중 하나의 기초번호 범위 안에 여러 개의 건물이 있을 경우, 기점방향의 하나는 기초번호가 부여되지만 그 이후에는 가지번호가 부여되기 때문이었다. <그림 25>에서 보는 것과 같이 기초번호 도로 세그먼트 부근에 존재하는 주소는 매칭되지 않는 결과가 나타났다.

이를 위해 가지번호 도로 세그먼트가 없는 주소를 정합시키기 위한 방안이 필요할 것으로 판단되었으며 또한 가지번호의 매칭을 위해서는 주소의 형태를 구분하는 것이 필요하게 되었다. 이를 위해 가지번호 도로 세그먼트를 통해 매칭이 가능한 주소는 가지번호1로 지정하고, 그렇지 않은 주소에 대해서는 가지번호2로 지정하고 정합률 및 위치정확도에서도 구분하여 실험을 진행하였다.



<그림 25> 가지번호 중 가지번호 도로 세그먼트로 매칭되지 않는 경우

2. 가지번호(일부)

위에서 가지번호의 형태를 2가지로 구분하였기 때문에 가지번호 도로 세그먼트가 존재하지 않는 주소데이터에 대해서는 가지번호2로 지정하였다.

가지번호2의 경우 수치지형도에서 도로중심선 데이터가 존재하지 않고, 주소가 부여되는 일관적인 규칙은 없지만, 지오코딩 시 어떤 위치에는 할당되어야 한다. 그러므로 본 연구에서는 해당 기초번호에 할당시킴으로써 정합률을 높이고자 하였다.

가지번호2의 매칭을 위해서는 가지번호1과는 다른 주소형식 변환이 필요하다. 즉, 가지번호2의 경우 ‘53 1 동선길’을 ‘1 동선길53’으로 변환한 반면, ‘53 동선길’로 형식을 변환한다. 즉 가지번호임에도 불구하고 가지번호는 삭제하고 지오코딩을 실시하는 것이다. 다만, 가지번호의 정보는 그대로 유지하되, 지오코딩을 위한 주소형식만 변환하였다.

기초번호 도로 세그먼트를 통해 지오코딩을 실시하였다. 지오코딩 결과, 57개의 가지번호2 자료에 대해 57개가 매칭되었다(표 11).

[표 11] 가지번호2 : 정합률

구분	가지번호2
주소 개수	57
매칭	57 (100)
매칭 안됨	0 (0)

실험 전의 기초번호를 기반으로 구축한 도로세그먼트는 전체 주소 중 기초번호 주소만을 매칭 되었지만, 가지번호의 정합률 개선을 위해 각 주소의 특징에 맞도록 실험을 진행하여 모든 주소를 지오코딩 할 수 있었다. 즉 정합률 개선 실험은 기초번호에 대해서는 기초번호의 도로 세그먼트를 통해 지오코딩 하였으며, 가지번호1은 가지번호 도로 세그먼트를 구축하고 주소형식을 이에 맞게 변환시켜 지오코딩을 실시하였다. 마지막으로 가지번호2는 기초번호 도로 세그먼트를 이용해 해당 기초번호로 지오코딩하여, 모든 주소를 점 사상으로 표현하였다.

제2절 위치정확도 개선 실험

위의 정합률 개선 실험을 통해 모든 주소가 지오코딩 되었다. 여기에서는 지오코딩된 주소에 대해 위치정확도를 평가한다. 위치정확도 평가는 정합률을 통해 구분한 도로 세그먼트에 따라 기초번호 및 가지번호2, 가지번호1의 순서로 진행하였다.

1. 기초번호 및 가지번호2

1.1 도로 세그먼트 구축 및 지오코딩

기초번호와 가지번호2의 도로 세그먼트는 도로중심선을 기초로 구축하였다. 이를 기초로 구축된 도로 세그먼트는 각 도로 간 연결부분이 모두 연결되어 있

지만, 건물 주소를 할당하기 위해서는 각 도로가 분리시켜 도로명을 부여하고 범위를 지정한다. 이러한 주소와 도로중심선의 시작 위치 차이는 위치정확도를 낮추는 요인으로 작용하고 있다.

위의 사항을 고려하여 도로중심선을 건물 주소의 시작위치에 따라 수정하여 도로 세그먼트를 구축하였다. <그림 26>은 위의 기초번호 위치정확도가 낮게 나왔던 지역으로써 동선길 부근 교차로의 원래의 도로 세그먼트와 수정된 도로 세그먼트를 비교하여 보여주고 있다. 동선길의 경우 해당 교차로를 통과하면서 주소가 부여되기 때문에 연결되어있는 반면, 동선5길(위쪽)과 정든길(왼쪽) 부근은 분리되어 구축된 것을 확인할 수 있다. 또한 단순히 교차로 중심부에서 해당 도로 세그먼트를 구분한 것이 아니라, 실제 주소가 시작되는 부분에서 도로 세그먼트가 시작하도록 구축하였다.

가지번호2는 기초번호 도로 세그먼트를 지오코딩되기 때문에, 위의 수정사항은 가지번호2의 지오코딩 시에도 적용된다.



<그림 26> 기초번호 및 가지번호2 : 도로 세그먼트 비교

기초번호는 주소부여방식이 기우식이므로 address locator는 US Streets로 설정하였고, 가지번호2는 일정 범위 내에서 순차적으로 주소가 부여되기 때문에 One Range로 지정하였다. 지오코딩 옵션은 두 주소 모두 철자민감도는 100으로 설정하였고, 기초번호에 대해서만 side offset을 1m로 설정하였다.

1.2 품질평가

기초번호에 대해 수정된 도로 세그먼트로 지오코딩 한 결과는 다음과 같다. 전체 589개 중에서 156개만이 오차가 존재하였으며 평균이 3.94m, 표준편차 2.64, 최소값 0.33, 최대값 14.91이다(표 12). 평균 및 표준편차를 비롯하여 모든 통계량에서 확연히 위치오차가 줄어들었다.

<그림 27>은 지오코딩된 기초번호의 위치정확도를 나타낸 것으로 전반적으로 위치정확도가 향상된 것을 확인할 수 있다. 특히 원본 도로 세그먼트로 지오코딩을 실시했을 때 위치오차가 크게 나타났던 지역의 위치정확도가 많이 향상되었다.

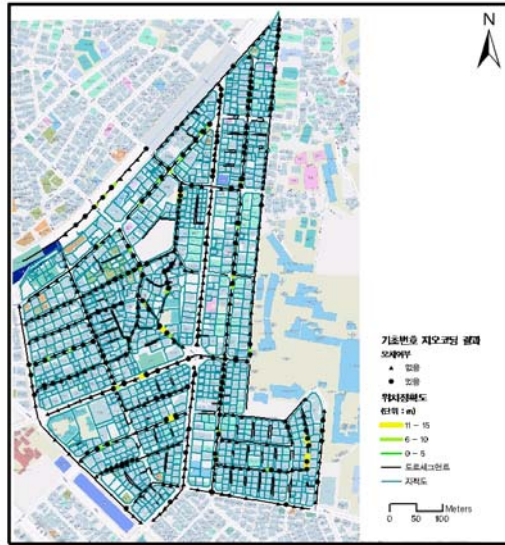
[표 12] 기초번호 : 위치오차 통계량

평균	3.94
표준편차	2.64
중앙값	3.27
최빈값	2.85
최소값	0.33
최대값	14.91
범위	14.58

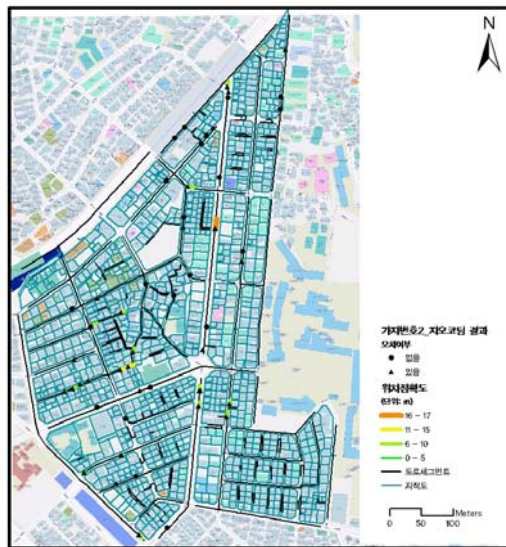
가지번호2는 기초번호 도로 세그먼트로 지오코딩하기 때문에 수정된 기초번호 도로 세그먼트로 지오코딩하여 위치정확도를 평가하였다. 57개의 데이터 중 36개의 데이터에서 오차가 발생하였으며 평균적으로 4.91m에 표준편차가 3.93, 최소값 0.05, 최대값 16.86, 최빈값 0.93이다(표 13). 즉 <그림 33>과 같이 가지번호2의 경우 표준편차가 다른 주소에 비해 크게 나온 것으로 보아 오차의 크기가 상당히 다양하게 나왔으나 최빈값이 0.93에서 알 수 있듯이 위치오차는 크지 않은 것을 확인하였다. 가지번호2의 오차는 가지번호와 가깝게 위치하고 있는 경우 오차가 적게 나타나는 반면 기초번호와 다소 멀리 위치하고 있는 경우 위치오차가 크게 나타나는 것으로 확인되었다.

[표 13] 가지번호2 : 위치오차 통계량

평균	4.91
표준편차	3.93
중앙값	4.05
최빈값	0.93
최소값	0.05
최대값	16.86
범위	16.81



〈그림 27〉 기초번호 : 위치정확도



〈그림 28〉 가지번호2 : 위치정확도

2. 가지번호1

2.1 도로 세그먼트 구축 및 지오코딩

가지번호1을 위해 가지번호 도로 세그먼트로 구축하였다. 이 때 가지번호는 기초번호와 달리 일정 범위에 대해 순차적으로 주소가 할당되기 때문에 기초번호와는 다른 방법의 도로 세그먼트 구축이 필요하다. 즉, 가지번호의 도로 세그먼트를 도로중심선을 기초로 구축하였을 경우 <그림 29>와과 같이 위치오차가 발생할 것이다. 그러므로 <그림 30>과 같이 도로 세그먼트를 구축하고, 기초번호와 마찬가지로 실제 주소가 시작되는 부분부터 끝나는 부분까지만 도로 세그먼트를 구축하는 전략이 필요하다. 이를 통해 실제로 구축한 도로 세그먼트가 <그림 31>이다.

주소의 형태에 따라 도로 세그먼트를 구축하였으며, 이에 대해 본 연구에서는 기존의 도로 세그먼트를 원본 도로 세그먼트라고 하고, 수정된 도로 세그먼트를 수정(본) 도로 세그먼트로 지정하여 사용하였으며, <그림 32>은 두 도로 세그먼트를 비교할 수 있도록 그림으로 나타내었다.

가지번호1의 address locator는 One Range로 설정하였다. 가지번호1의 경우 도로세그먼트 구축 시 확인하였듯이 기점으로부터 종점까지 연결된 주소지를 따라 연결되어 때문이다. 지오코딩 옵션은 철자민감도만 100으로 설정하였으며, side offset 및 end offset은 One Range에서는 지원하지 않아 지정하지 않았다.



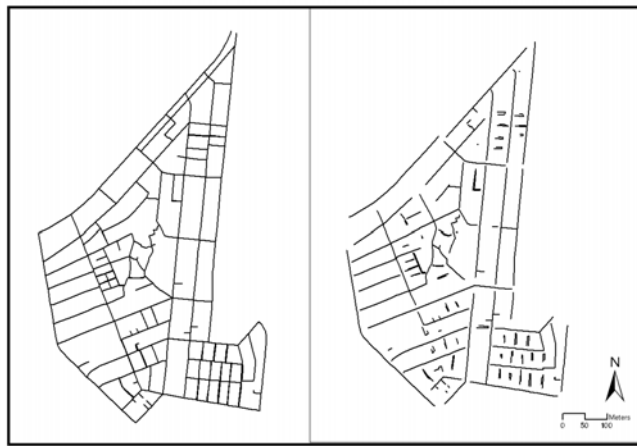
〈그림 29〉 가지번호1 : 도로중심선에 기초한 도로 세그먼트 구축 예



〈그림 30〉 가지번호1 : 도로중심선을 수정한 도로 세그먼트 구축 예



〈그림 31〉 가지번호1 : 도로 세그먼트 비교



〈그림 32〉 도로 세그먼트 비교(전체 도로망)

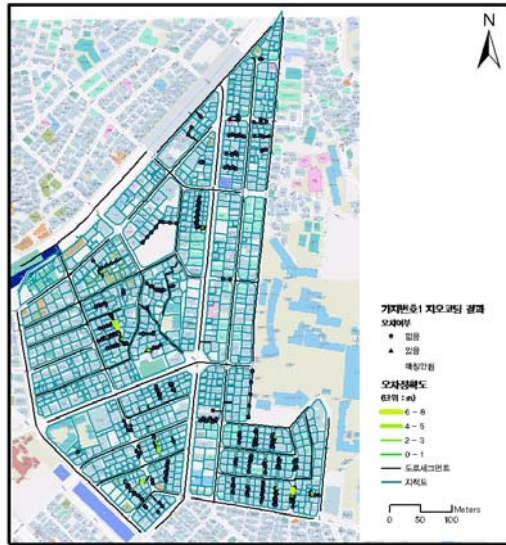
1.2 품질평가

가지번호1은 위치정확도 개선을 위해 주소의 특징에 맞도록 도로 세그먼트를

수정하였다. 수정된 도로 세그먼트로 지오코딩한 결과, 전체 266개의 주소 중 24개에 대해서만 위치오차가 발생하였다. 오차의 통계량을 보면, 평균 위치정도가 2.43m, 표준편차 1.72, 최소값 0.24, 최대값 7.92로써 위치정확도가 상당히 높게 나왔다(표 14). <그림 33>에서도 보듯이 위치오차가 거의 없고 지도의 중앙부인 하나로3거리 22 부근에서 위치오차가 높게 발생하였다. 이 지역은 건물의 크기가 일정하지 않은 곡선지역으로 오차가 높게 나타났다.

[표 14] 가지번호1 : 위치오차 통계량

평균	2.43
표준편차	1.72
중앙값	2.14
최빈값	2.14
최소값	0.24
최대값	7.92
범위	7.68



〈그림 33〉 기초번호1 : 위치정확도

제3절 결과 분석

1. 정합률

여기에서는 위의 지오코딩된 결과를 종합적으로 분석하고자 한다. 먼저 정합률을 보면, 원본 도로 세그먼트를 통해 지오코딩하였을 때 기초번호인 589개만이 정합되었다. 이를 개선하기 위해 주소를 형태별로 구분하여 도로 세그먼트를 구축하거나 주소 형식을 변환하여 지오코딩하였다. 이 결과 [표 15]와 같이 정합률이 개선되었다.

[표 15] 주소형태에 따른 정합률 비교

구 분	원본		수정본			
	전체	기초번호	전체	기초번호	가지번호	
					가지번호1	가지번호2
주소 개수	912	589	912	589	266	57
매칭	589 (64.59)	589 (100)	912 (100)	589 (100)	266 (100)	57 (100)
매칭 안됨	323 (35.41)	0 (0)	0 (0)	0 (0)	0 (0)	0 (0)

2. 위치정확도

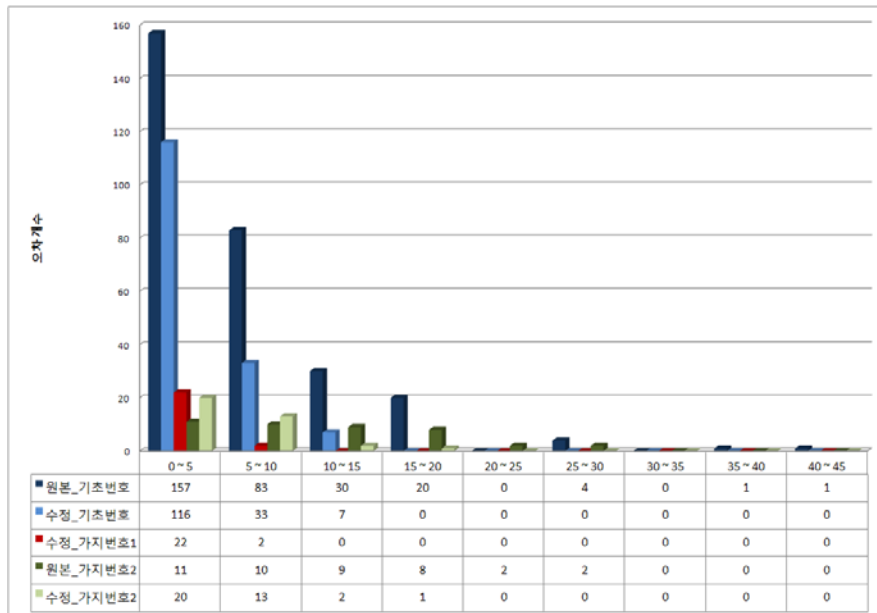
정합률과 마찬가지로 위치정확도도 주소의 형태별로 나누어 살펴보도록 한다. 기초번호의 경우 오차의 수와 오차 통계량에서 모두 개선되었다. [표 16]과 <그림 34>와 같이 오차 발생 수가 원본의 경우 311개에서 수정본은 156로 급격히 감소하였으며, 오차의 크기도 감소하였음을 알 수 있다. 원본의 경우 35~40, 40~45 구간에 존재하였던 오차가 수정본에서는 나타나지 않았다.

[표 16] 전체 주소 : 도로 세그먼트 간 위치오차 통계량 비교

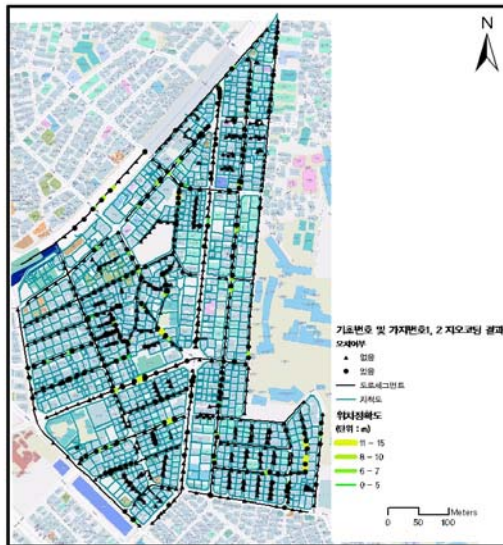
구분		오차 수	평균	표준편차	최소값	최대값	범위
기초번호	원본	311	6.88	6.12	0.10	40.81	40.71
	수정	156	3.94	2.64	0.33	14.91	14.58
가지번호1	수정	36	4.91	3.93	0.05	16.86	16.81
가지번호2	원본	20	16.72	4.96	10.92	28.17	17.25
	수정	36	4.91	3.93	0.05	16.86	16.81

가지번호1의 경우, 원본에서 매칭되지 않았기 때문에 원본과의 비교가 불가능하여 수정 도로 세그먼트에 대해 오차를 나타내었다. 본 연구에서 위치정확도 개선을 위한 실험에서 가지번호2는 원본 도로 세그먼트에서 매칭시키지는 않았으나, 원본 도로 세그먼트에서 매칭되었을 때와 수정 도로 세그먼트에서 매칭되었을 때 위치정확도 비교를 위해 매칭시켜보았다. 매칭 시킨 결과 오차의 발생 수는 원본보다 수정본에서 오히려 많이 나타났으나 오차의 크기가 줄어들면서 위치정확도가 개선되었다.

<그림 35>과 같이 품질개선 실험결과 전체 주소들은 오차의 정도가 확연히 감소한 것을 확인할 수 있다. 오차가 다소 크게 나타나는 지역은 해당 도로 구간에서 어떤 건물이 다른 건물에 비해 면적이 넓기 때문에 나타나는 오차이므로, 대부분의 오차는 이번 실험을 통해 개선되었다고 할 수 있다.



〈그림 34〉 전체주소 위치정확도 비교



〈그림 35〉 전체 주소 : 수정 도로 세그먼트의 위치정확도

제5장 결론

본 연구는 새주소체계에서 지오코딩에 대한 방법 및 과정을 고찰하고, 실험지역에 지오코딩을 실시하고 품질평가를 통해 보다 개선된 품질의 지오코딩의 방법을 제시하고자 하였다.

본 연구를 통해 얻은 결과를 종합하면 다음과 같다.

첫째, 새주소체계의 변화에 따라 기존에 적용할 수 없었던 도로망기반의 주소체계를 이용한 지오코딩이 가능해졌다. 필지기반의 경우 지적관련 데이터의 접근성이 낮아 지오코딩이 어려웠던 반면, 도로망기반 지오코딩은 데이터의 취득이 용이하기 때문에 사용자나 연구자들이 지오코딩된 주소를 이용하여 보다 효율적으로 공간분석을 할 수 있다.

둘째, 주소를 특징에 따라 세분화하여 지오코딩을 실시하면 정합률을 향상시킬 수 있다. 기초번호를 중심으로 구축한 도로 세그먼트는 기초번호에만 적합하여 가지번호는 매칭시킬 수 없었다. 그러므로 새주소체계를 해당 도로 세그먼트의 특징에 따라 3가지의 형태 즉, 기초번호와 가지번호1, 가지번호2로 구분하여 지오코딩을 실시하였으며 모든 주소가 매칭되었다.

셋째, 도로 세그먼트 구축 시 주소의 특징에 따라 도로중심선을 수정함으로써 위치정확도를 증가시킬 수 있다. 도로중심선은 모든 도로를 따라 연결되도록 구

측되어 있지만, 주소는 도로를 구분하여 주소가 할당되므로, 이러한 부분에서 위치오차가 많이 발생하였다. 그러므로 해당 주소가 시작되는 부근부터 도로 세그먼트를 구축하면 위치오차를 줄일 수 있다.

본 연구에서의 한계점은 다음과 같다.

본 연구에서는 가지번호2에 대해 정합률을 개선시키는데 초점을 두었고, 부차적으로 수정된 기초번호 도로세그먼트를 이용해 위치정확도를 개선시켰다. 그러므로 가지번호2의 위치정확도를 개선시킬 수 있는 궁극적인 연구가 필요할 것으로 보인다.

또한 실험지역을 도시지역으로 한정지음으로써 농촌지역에서의 지오코딩 품질을 평가할 수 없었다. 농촌지역은 거주지가 밀집되어 있지 않기 때문에 도시지역에 비해 지오코딩의 품질향상을 위해 본 연구에서 제시한 방안과는 다른 방안이 필요할 것이다.

한편, 주소는 기초번호 및 가지번호로만 구성되는 것이 아니라 연립주택 및 아파트의 경우 상세주소도 주소에 포함되는데, 본 연구에서는 상세주소는 고려하지 않았기 때문에 위의 요소를 고려한 지오코딩 방법론이 필요할 것으로 보인다.

지도는 위치에 대한 정보를 나타내는 가장 보편적이고, 많은 양의 정보를 가장 명료하게 나타내는 수단이다. 정보화 사회에서 지도는 GIS를 통해 여러 정보들과 조합되고 분석되어 의사결정 과정에 참여한다. GIS에서 지도가 사용되기 위해서는 GIS에서 사용할 수 있는 데이터로 구축하는 과정이 필수적이다.

특히, 점 사상의 경우 데이터 구축을 위해 많은 시간과 참조자료가 필요하다. 지오코딩은 점 사상의 데이터 구축에 사용되어 왔으며, 기존의 필지기반의 지오코딩에 비해 도로망기반 지오코딩이 보다 효과적으로 데이터 구축에 사용될 수 있다. 그러나 이에 대한 연구는 초기단계로 활성화되어 있지 않으므로, 보다 체계적이고 지속적으로 이에 대한 연구가 이루어져야 할 것이다.

참 고 문 헌

1. 국내 문헌

- 강영옥, 이동연, 1996, 위치정보제고를 위한 주소표시제도 개선방안, 국토계획, 31(6), 227-247.
- 구은훈, 2005, Web-GIS를 이용한 새주소체계 활용방안에 관한 연구(경산시를 중심으로), 대구대학교 행정대학원 석사학위 논문.
- 김길용, 2008, 새주소 현장관리를 위한 Mobile GIS 시스템 개발, 전북대학교 산업기술대학원 석사학위논문.
- 손 철, 2002, Do Different GIS-based Geocoding Methods Result in Different Parameter Estimates for Location Variables from a Hedonic Housing Price Model?, 국토계획, 37(1), 153-168.
- 연상호, 유준상, 2001, 제천시 새주소 관리 및 안내 시스템 구축에 관한 연구, 한국지리정보학회지, 4(3), 13-20.
- 오충원, 2007, 인터넷 환경에서 지도의 융합에 관한 연구, 대한지리학회 연례 학술대회 발표집, 315-318.
- 유상철, 2008, 새주소의 생활화 · 조기정착방안에 관한 연구, 건국대학교 부동산대학원 석사학위논문.
- 이기환, 2004, 새주소 부여 사업의 유지관리 개선방안에 관한 연구 : 춘천시 사례를 중심으로, 강원대학교 정보과학대학원 석사학위 논문.
- 이민영, 오민수, 이지영, 신영철, 2001, 새주소 관리시스템에 관한 연구-청주

- 시 도로명및 건물번호 부여체계, 한국정보지리학회지, 4(1), 47-56.
- 이지영, 김학열, 2006, A Geocoding Method Implemented for Hierarchical Areal Addressing System in Korea, 한국GIS학회지, 14(4), 403-419.
- 이진하, 2007, 지번과 새주소의 효율적인 발전방안, 명지대학교 산업대학원 석사학위논문.
- 임재일, 2001, 도로명과 건물번호에 의한 새주소부여사업의 효율성 개선방안에 관한 연구, 부산대학교 환경대학원 석사학위논문.
- 정재준, 1998, 사용자 측면을 고려한 수치지도의 등급분류에 관한 연구, 서울대학교 대학원 석사학위논문.
- 지종덕, 2005, 한국 새주소체계의 발전방향에 관한 연구, 한국지적학회지, 21(2), 37-52.
- 행정자치부, 2007, 묻지않고 찾아가는, 자신있는 발걸음(보도자료).
- 문화관광부 국어심의회, 1997, 「도로명 부여에 관한 용어」.

2. 국외 문헌

- Block, R. 1995, Geocoding of crime incidents using the 1990, TIGER file: the Chicago example, In C. R. Block, M. Dabdoub, and S. Fregly, eds., Crime analysis through computer mapping, Washington, D.C. : *Police Executive Research Forum*, 15.
- Bonner, M. R., Han, D., Nie, J., Rogerson, P., Vena, J. E., and

- Freudenheim, J. L., 2003, Positional accuracy of geocoded addresses in epidemiologic research, *Epidemiology*, 14(4), 408–411.
- Boulos, M. N. K., 2004, Towards Evidence-Based, GIS-Driven National Spatial Health Information Infrastructure and Surveillance Services in the United Kingdom, *International Journal of Health Geographics*, 3(1).
- Cayo, M. R. and T. O. Talbot, 2003, Positional Error in Automated Geocoding of Residential Addresses, *International Journal of Health Geographics*, 2(10).
- Chung, K., Yang, D. H., and Bell, R., 2004, Health and GIS : toward spatial statistical analyses, *Journal of Medical Systems*, 28(4), 349–60.
- Clapp, J. M., Mauricio, R. and Grant, T., 1997, How GIS can Put Urban Economic Analysis on the Map?, *Journal of Housing Economics*, 6, 368–386.
- Goldberg, D. W., Wilson, J. P. and Knoblock, C. A., 2007, From Text to Geographic Coordinates : The Current State of Geocoding, *Journal of the Urban and Regional Information Systems Association*, 19(1), 33–46.
- Dearwent, S. M., Jacobs, R. R. and Halbert J. B., 2001, Locational uncertainty in georeferencing public health datasets, *Journal of Exposure Science and Environmental Epidemiology*, 11,

329–334.

- Drummond, W. J. 1995, Address matching : GIS technology for mapping human activity patterns, *Journal of the American Planning Association*, 61(2), 240–51.
- Evans, G. W. and Kantrowitz, E., 2002, SOCIOECONOMIC STATUS AND HEALTH : The Potential Role of Environmental Risk Exposure, *Annual Review of Public Health*, 23, 303–331.
- Gatrell, A. C., 1989, On the spatial representation and accuracy of address-based data in the United Kingdom, *International Journal of Geographical Information Systems*, 3(4), 335–48.
- Geronimus, A. T., Bound, T. J., and Neidert L. J., 1995, On the validity of using census geocode characteristics to proxy individual socioeconomic characteristics, Technical Working Paper 189, Cambridge, MA : *National Bureau of Economic Research*.
- Karimi, H. A., M. Durcik, and W. Rasdorf, 2004, Evaluation of uncertainties associated with geocoding techniques, *Journal of Computer-Aided Civil and Infrastructure Engineering*, 19(3), 170–185.
- Kravets N. and Hadden, W. C., 2007, The accuracy of address coding and the effects of coding errors, *Health place*, 13, 293–298.
- Krieger, N., 1992, Overcoming the absence of socioeconomic data in medical records : validation and application of a census-based methodology, *American Journal of Public Health*, 82(5),

703–710.

- Krieger, N., Waterman, P. D., Chen, J. T., Soobader, M. J., and Subramanian, S. V., 2003, Monitoring socioeconomic inequalities in sexually transmitted infections, tuberculosis, and violence : geocoding and choice of area-based socioeconomic measures, *Public Health Reports*, 118(3), 240–60.
- Nicoara, G., 2005, *Exploring the Geocoding Process : A Municipal Case Study using Crime Data*, Masters thesis, The University of Texas at Dallas, Dallas, Tx, USA.
- Olligschlaeger, A. M., 1998, Artificial Neural Networks and Crime Mapping, In D. Weisburd and T. McEwen (Eds.), *Crime Mapping and Crime Prevention*, Volume 8 of Crime Prevention Studies (Monsey, NY USA : Criminal Justice Press), 313–347.
- Paul, A. Z., 2008, A comparison of address point, parcel and street geocoding techniques, *Computer, Environment and Urban systems*, 32, 214–232.
- Ratcliffe, J. H., 2004, Geocoding crime and a first estimate of a minimum acceptable hit rate, *International Journal of Geographical Information Science*, 18(1), 61–72.
- Ratcliffe, J. H., 2001, On the Accuracy of TIGER-Type Geocoded Address Data in Relation to Cadastral and Census Areal Units, *International Journal of Geographical Information Science*, 15(5), 473–485.

- Robert S. A., 1999, SOCIOECONOMIC POSITION AND HEALTH : The Independent Contribution of Community Socioeconomic Context, *Annual Review of Sociology*, 25, 489–516.
- Rushton, G., Armstrong, M., Gittler, J., Greene, B., Pavlik, C., West, M., and Zimmerman, D., 2006, Geocoding in cancer research—review, *American Journal of Preventive Medicine*, 30(2), S16–S24.
- Thakkar, S., Knoblock, C. A., Ambite, J. L., Shahabi, C., 2002, Dynamically composing web services from on–line sources, *Proceeding of American Association for Artificial Inteligence*.
- Vine, M. F., Degnan, D., and Hanchette, C., 1998, Geographic information systems : their use in environmental epidemiologic research, *Journal of Environmental Health*, 61, 7–16.
- Whitsel, E. A., Rose, K. M., Wood, J. L., Henley, A. C., Liao, D. and Heiss, G., 2004, Accuracy and repeatability of commercial geocoding, *American Journal Epidemiology*, 160, 1023–1029.
- Wu, J., Funk, T. H., Lurmann, F. W., and Winer, A. M., 2005, Improving Spatial Accuracy of Roadway Networks and Geocoded Addresses, *Transactions in GIS*, 9(4), 585–601.
- Zimmerman, D. L., Fang, x., Mazumdar, S. and Rushton, G., 2007, Modeling the probability distribution of positional errors incurred by residential address geocoding, *International Journal of Health Geographics*, 6(1).

3. 인터넷자료

ESRI, <http://www.esri.com>

새주소안내시스템 <http://www.juso.go.kr>

ABSTRACT

A study on the geocoding implementation of
Korea's new address system

Roh, Hye jung

Department of Geography

Graduate school

Sungshin Women's University

(Supervised by Professor Jeong, Jae-joon)

Addresses are one of conceptual means for location in modern world. They have important information of personal behavior or spatial analysis. Recently, address system in Korea is changing from parcel based system to street based address system. Parcel based system has been no longer providing of location information due to rapid industrialization, increase of population or sprawl of city. Therefore new address system based on street network introduced.

The objectives of this study are improved the geocoding quality, matching rate and positional error. This study reviews the fundamental geocoding process, concept, and address system in Korea.

Dongseon-dong, Seongbuk-gu, Seoul is selected as study area.

Geocoding is the process of assigning an x, y coordinate value to the description of a place by comparing the descriptive location-specific elements to those in the reference data.

The main results of this study were summed up as followings;

Firstly, the geocoding was possible as new address system based street network. Secondly, matching rate in geocoding quality was improved by applying some matching method by characteristics of addresses. Thirdly, positional errors in geocoding quality were reduced by modifying geocoding base file.

In this study, solutions for the geocoding errors were presented and geocoding base file modifications were made to accommodate address-specific needs.