

지역분석을 위한 웹 기반 통계GIS 연구⁺

박기호* · 이양원**

A Study on the Statistical GIS for Regional Analysis

Key-Ho PARK · Yang-Won LEE

요 약

우리 나라 국가지정통계자료는 대부분 행정구역을 기초단위로 하는 지리/지역정보로서 계량적 지역분석기법을 통해서 그 가치가 더해진다. 본 연구는 지역분석을 위한 통계정보 및 통계지도 서비스 시스템의 한 선도적 모델을 제시하는데 그 목적을 둔다. 본 연구에서는 서울시를 사례로 하여 인구주택 총조사 및 사업체기초통계 원시자료 등을 재 집계하여 지역통계 데이터베이스를 구축하였다. 통계GIS의 핵심 지역분석모듈은 사회경제 자료분석에서 활용의 빈도가 높은 지역구조 및 격차 분석, 공간 분산분석, 상관분석, 회귀 및 잔차분석, 공간적 자기상관분석 등으로 구성되었다. 시스템 설계는 지역통계 데이터베이스와 지역분석기법 모듈, 그리고 통계지도를 웹 환경에서 통합하기 위해 어플리케이션 서버를 포함한 다양한 정보기반기술을 채택하여 이루어졌다. 지역분석을 위한 입출력 과정에서 대화형 통계지도가 사용된다. 본 연구를 통해 제안된 통계GIS의 특징은 미리 저장된 데이터와 분석기능이 사용자와의 양방향 웹 인터페이스를 통해 제한 없이 확장될 수 있다는 점이다. 사용자의 요구에 따라 분석변수의 자유로운 변형과 합성이 연산자 구문해석기를 통해 가능하고, 이를 바탕으로 새로운 지역분석 알고리즘의 구현이 실시간으로 가능하다. 또한 데이터 측면에서는 사용자가 업로드한 자료를 잠정적으로 시스템에 통합하여 확장된 가상 데이터베이스를 만들고 이에 대해 여러 분석을 수행할 수 있다. 웹 기반 통계GIS 모델의 초기평가를 통해 각종 지역통계자료로부터 주문형 통계지도의 작성과 유통이 동시에 가능함이 확인되었다.

주요어 : 지역분석, 통계지도, 인터넷 지리정보시스템

+ 본 연구는 2001년도 서울대학교 대학연구센터 연구비 지원을 받았습니다.

* 서울대학교 지리학과 부교수 (Associate Professor, Department of Geography, Seoul National University)

** 서울대학교 지리학과 박사과정 (Ph. D Student, Department of Geography, Seoul National University)

ABSTRACT: A large suite of official statistical data sets has been compiled for geographical units under the national directives, and it is the quantitative regional analysis procedures that could add values to them. This paper reports our attempts at prototyping a statistical GIS which is capable of serving over the Web a variety of regional analysis routines as well as value-added statistics and maps. A pilot database of some major statistical data was ingested for the city of Seoul. The baseline subset of regional analysis methods of practical usage was selected and accommodated into the business logic of the target system, which ranges from descriptive statistics, regional structure/inequality measures, spatial ANOVA, spatial (auto)correlation to regression and residual analysis. The leading-edge information technologies including the application server were adopted in the system design and implementation so that the database, analysis modules and analytic mapping components may cooperate seamlessly behind the Web front-end. The prototyped system supports tables, maps, and files of downloadable format for input and output of the analyses. One of the most salient features of our proposed system is that both the database and analysis modules are extensible via the bi-directional interface for end users; The system provides users with operators and parsers for algebraic formulae such that the stored statistical variables may be transformed and combined into the newly-derived set of variables. This functionality eventually leads to on-the-fly fabrication of user-defined regional analysis algorithms. The stored dataset may also be temporarily augmented by user-uploaded dataset; The extension of this form, in essence, results in a virtual database which awaits for users commands as usual. An initial evaluation of the proposed system confirms that the issues involving the usage and dissemination of information can be addressed with success.

Keywords : Regional Analysis, Statistical Map, Web GIS

1. 서 론

1.1. 연구 배경

현재 국내 여러 기관에서는 147여종에 이르는 국가지정통계정보를 막대한 재원과 노력을 기울여 조사하고 있다. 최근

국가 통계자료의 활용도 제고를 위한 관심이 모아지고 있으며 김영표와 한선희(1999)는 행정업무는 물론 학계와 일반인의 통계정보 활용에 큰 도움을 주기 위해 통계자료가 표나 그래프 형태와 더불어 지도로 가공된 국가통계지도의 제작이 필요하다고 지적하고 있다.

이러한 통계지도가 담아내야 할 정보의

내용은 그 폭과 깊이가 수요자에 따라 매우 다양할 것이다. 통계지도는 국가GIS 사업으로 제작되고 있는 기본도나 공통 주제도와는 달리 그 내용에 있어서 합의가 이루어지기 힘들고, 또한 그 활용빈도에 따라 공통분모가 되는 일련의 표준 통계지도가 설정된다 하더라도 언제나 다른 형태의 통계정보와 요약이 필요하기 마련이다. 따라서 통계지도는 GIS의 일부 핵심기능을 갖추고 수요자의 필요에 따라 대화형 혹은 주문형으로 작성하고 변형할 수 있는 형태가 바람직하다.

정책통계자료의 80% 이상이 지역별 비교가 필요한 지리적 정보로 작성되고 있다는 사실은 주기적인 지역분석관련 통계의 재창출이 기초통계자료의 매우 중요한 목적임을 시사한다. 지역분석 (Regional Analysis)은 매우 포괄적인 개념으로서 지역의 경제, 사회, 인구, 혹은 시설 분포의 구조적 특성을 파악하거나 지역간의 비교, 그리고 지역의 특정 현상을 이해하고 설명하려는 모형의 정립 등 정량적 방법론에 그 기초를 두고 있다. 통계정보와 통계지도는 그 자체가 목적이 아니라 지역분석 절차나 기법을 통해 그 의미가 부여된다고 해도 과언이 아니다. 국가통계정보의 활용도 제고를 위해서는 통계지도와 지역분석에 있어서 그 가치가 검증된 계량기법과 긴밀하게 연계하는 방안의 모색이 시급하다.

한편 통계정보 및 통계지도의 제작과 유통체제는 누구나 언제 어디서나 접근과 취득이 쉬운 인터넷 서비스 모델에 기반을 두어야 할 것이다. 최근 폭발적이라 할 수 있는 웹(World Wide Web)의 접근

인프라와 다양한 콘텐츠의 구축은 일상생활과 관련된 영역뿐만 아니라 학술연구의 보조수단으로까지 그 정보제공의 영역이 확대되고 있고, 웹을 통해 연구에 필요한 참고문헌을 확보하거나 통계자료를 입수하는 것은 매우 보편적인 일이 되어가고 있다. 그러나 아직까지는 웹 서비스가 자료원으로서의 역할에 머물러 있으며, 자료로부터 고급정보를 추출하는 분석도구로서의 역할은 매우 미흡한 실정이다. 국내의 KOSIS와 Stat-Korea와 같은 기존의 웹 혹은 온라인 통계 시스템은 통계자료를 체계적으로 분류하여 제공하기는 하지만, 이 자료에 대한 기초통계분석도 자료의 다운로드가 이루어진 후 개별 사용자의 데스크 탑에서 오프라인으로 이루어지고 있다.

최근 활발히 구축되고 있는 웹 GIS는 지도정보의 온라인 분석기능이라는 측면에 있어서 상당한 성과를 거두고 있다 (Plewe, 1997; Harder, 1998). 즉 단순한 자료의 조회를 넘어 최단경로 선정이나 대화형 단계구분도 제작, 그리고 지도대수 등의 연산에 이르는 프로세싱 서비스가 가능하게 되었다(박기호, 1997). 하지만 이러한 온라인 처리시스템에서 아직 미흡한 기능 중에 하나는 사용자의 자료를 처리해 주는 양방향 서비스라 하겠다. 현재 웹 GIS 사이트의 운영 양식은 자체 콘텐츠 서버에 저장된 자료를 사용자의 요구에 맞게 처리하여 제공하는데 불과하다. 이것은 고객의 필름을 받아 현상 인화해 주지 못하고 미리 준비된 포스터만 제작하는 사진관과 크게 다르지 않다.

GIS 분야에서는 지역분석과 관련한 활용

이나 공간분석이나 모델링 기법의 접목은 그 동안 꾸준한 논의가 이루어지고 있다 (Fischer et al., 1996; Chou, 1997; Fotheringham and Wegener, 2000). 공간자료의 통계적 처리와 분석에 있어서도 GIS와 통계모델의 연동을 위한 시도들이 이루어지고 있고 (Lee and Wong, 2001), 아울러 특정 분야에 전문화된 분석모델의 통합에 관한 연구도 보고되었다(박기호, 2000). 그러나 이러한 시도들은 웹 환경을 염두에 두지 않고 Stand-Alone 방식의 범용 GIS의 틀에서 벗어나지 못했다.

지금까지 지적된 문제점들을 놓고 볼 때, 정보화 시대 그리고 지식기반사회의 가장 기본적인 사회 간접자본이라 할 수 있는 통계자료 및 지리정보의 활용과 유통에 대한 새로운 모델이 필요한 시점이다.

1.2. 연구 목적

본 연구에서는 행정정책업무에는 물론 지리학 등 학술연구에 필요한 주 자료인 행정구역 단위의 통계자료의 효과적인 유통과 활용에 있어서 전술한 문제점들을 극복하는 방안을 모색한다. 본 연구의 목적은 지역분석을 위한 통계정보 및 통계지도 서비스 시스템의 한 선도적 모델을 제시하고 그 특징점과 기술적 타당성을 시범시스템 구축을 통하여 평가하는 데 있다. 이를 위한 연구의 세부 항목과 범위는 다음과 같다.

첫째, 행정구역 단위의 기초통계자료를 바탕으로 이루어지는 지역분석의 계량적 방법론을 검토하여 통계GIS의 핵심 Business

Logic에 해당하는 분석모델을 제시한다. 분석모델의 범위는 기술적 통계량 산출에서부터, 지역구조분석, 지역격차 분석, 공간 분산분석 (Spatial ANOVA), 상관 분석, 회귀 분석, 공간적 자기상관 분석 등 사회경제적 자료분석에서 활발히 사용되고 그 가치가 검증된 대표적 공간통계 분석 기법 중에서 선별한다.

둘째, 지역분석을 위한 기초 지역통계 정보 데이터베이스를 마련한다. 방대한 범위와 용량의 인구, 경제 및 사회에 대한 조사자료를 행정구역별/항목별로 집계할 수 있는 데이터 스키마를 설계하고, 또한 원시자료 뿐만 아니고 사용자의 자료를 데이터베이스에 능동적으로 컴파일하기 위한 트리거 (Trigger)나 OLAP 등 데이터베이스 기술의 접목 방안을 모색한다.

셋째, 위에서 도출된 지역통계 데이터베이스와 지역분석기법 모델을 웹 환경에서 연계하여 다양한 공간통계정보와 통계지도를 제작하는 시스템의 설계 컨셉트를 제시한다. 목표 시스템의 설계는 유연성과 다기능성, 자료처리의 효율성 등에 초점을 두고 앞서 제기된 현행 시스템들의 문제점을 극복한다. 시스템 구축의 기술적 측면에서는 Web-DB 연동 시스템의 안정성과 효율성을 극대화 할 수 있는 소프트웨어 컴포넌트 연계 및 배치 (deploy) 모델을 관련 기반기술을 검토하여 대안을 마련한다. 이에 따라 종합된 시스템 아키텍처를 바탕으로 지역분석 통계GIS의 모형을 구현하고 평가한다.

2. 통계를 이용한 지역분석 방법론

지역분석은 Isard(1960)의 저술에서 그 학문적 태동을 찾는다. 지역분석은 지리학 경제학 등 사회과학 전반에 걸쳐서, 또한 도시 및 지역계획과 같은 응용학문 분야에서 유용한 계량적 분석기법으로 자리를 잡고 있다. 지역분석 기법은 매우 광범위하므로, 본 연구에서는 통계적 모형, 특히 가장 기본적인 요약 지수로 논의의 범위를 제한하였다. 이는 본 연구의 목표시스템을 시범구축하기에 적합하며 실질적인 지역분석의 주요부분을 수행하기에 부족함이 없는 범위이기 때문이다. 이 절에서 논의되는 기법들에 대한 자세한 내용은 지역분석의 계량기법을 다루는 문헌들(Krueckeberg and Silvers, 1974; 최재선, 1980; Massey and Denton, 1988; 이희연, 1990)에서 찾아볼 수 있다.

2.1. 기술적 통계와 변수의 변환 및 합성

지역분석을 포함한 모든 계량적 분석은 자료의 기본적인 기술 통계량(Descriptive Statistics)을 산출하는 것으로서 시작된다. 분석변수의 분포에 대하여 산술평균, 최빈치, 중위수 등으로 그 대표값을 산출하고, 또한 변량과 분산된 정도는 표준편차, 사분위편차, 그리고 분포의 특성은 왜곡도와 첨도와 같은 통계량으로 흔히 요약된다.

한편 지역분석을 위한 자료는 지리공간상에 분포되어 있는 현상에 대한 것이 대부분이다. 이렇게 공간성을 내포하고 있는 자료는 그 요약에 있어서 일반 통계와는 다른 처리가 요구된다. 예를 들어 단

위지역들의 인구나 면적의 차이를 반영하는 가중평균(Weighted Mean)과 같은 통계량이 지역의 분석에 있어서 더욱 의미 있는 대표값이 된다. 특정 변수의 인구가중평균 혹은 면적가중평균을 구하기 위해서는 변수의 변환(Transformation)이 필요하다. 변수 변환은 단일변수의 대수적 변환을 통해 이루어지며, 표준 정규값(Standard Z-Score), 로짓(Logit), Freeman-Tukey 변형 등이 그 대표적이 예라 할 수 있다.

변수는 서로 다른 변수들을 합성해서 유도되어지기도 한다. 간단한 예로 인구 P 와 면적 A 라는 두 변수에서 인구밀도 D ($=P/A$) 라는 새로운 합성변수가 만들어진다. 지역분석의 많은 자료는 녹지율, 남녀 성비, 주택보급율과 같이 변수간의 비율(Rate, Proportion, Percentage)로 합성되는 경우가 많다.

이처럼 기술통계량과 더불어 변수의 변형과 합성은 가장 기본적이면서 중요한 분석기법으로 이해해야 될 것이다. 지역의 현황과 특성을 기술해내는 사회지표들도 궁극적으로 변수의 변형과 합성의 한 형태로 보아도 무방하다. 이러한 의미에서 이들은 지역분석을 위한 통계GIS에 긴요한 모듈인 것이다.

2.2. 지역구조 분석

지역은 어떤 특수한 기능으로 전문화되어 가고 성장한다. 지역간 산업구조의 심화 또는 고도 전문화의 양상과 그 차이에 대한 분석은 지역연구에 있어서 많은 시사점을 제공한다. 지역구조 분석(Regional Structure Analysis)의 척도는 통상

적으로 산업구조의 분석에 있어서 그 가치가 검증된 지표들이며 각각의 계산식과 설명은 업종과 고용인력을 바탕으로 한다. 하지만 여기서 다루는 지역의 집중과 분화, 구조적 유사성과 변화 등은 산업과 고용인력의 측면 뿐 아니고 지역의 토지 이용, 인구유형, 편의시설 등 지역의 사회구조 제반에 적용할 수 있는 개념이며 척도이다. 지역구조 분석에 주로 이용되는 지표는 [표 1]과 같이 정리될 수 있으며, 이러한 지표들 중에서는 전체 대상지역을 요약하여 하나의 값으로 표현되는 것도 있고, 단위지역마다 값이 할당되어 지도화가 가능한 것도 있다.

입지계수(Location Quotient)는 특정산업의 지역별 분포를 상대적 척도로 비교하는 경우 유용한 계수이고, 집중계수(Coefficient of Localization)는 특정산업 k의

지역분포, 즉 산업입지의 공간적인 분산 정도를 일정 시점 또는 시점간 입지분포의 변동을 비교하는 척도이며, 특화계수(Coefficient of Specialization)는 각 단위지역에 업종이 얼마나 집중되고 전문화되어 있는가를 판별해내는 척도이다. 이 지표들의 수식에서 N은 전체지역의 총 사업체수, N^k 는 전체지역의 k 업종 사업체수, n_i 는 i지역의 총 사업체수, 그리고 n_i^k 는 i지역의 k 업종 사업체수를 의미한다.

입지연관계수(Coefficient of Geographical Association)는 한 업종이 특정지역에 입지하는데 있어서 타업종과의 연관도를 평가하는 척도로서 수식에서 k, l은 각 업종, i는 각 단위지역을 의미한다. 구조유사 척도(Measure of Structural Similarity)는 산업의 생산 또는 여타 사회구조에 있어 유사한 지역들을 판별하는 방법으로서 두 지

<표 1> 지역구조 분석 지표

지 표	공 식	비 고
입지계수	$LQ_i = \frac{n_i^k}{n_i} / \frac{N^k}{N}$	단위지역별로 값이 할당됨, 지도화 가능
집중계수	$C^k = \frac{1}{2} \sum_i \left \frac{n_i^k}{N^k} - \frac{n_i}{N} \right $	속성별로 단일 값을 가짐
특화계수	$C_i = \frac{1}{2} \sum_k \left \frac{n_i^k}{n_i} - \frac{N^k}{N} \right $	단위지역별로 값이 할당됨, 지도화 가능
입지연관 계수	$C^{k,l} = \frac{1}{2} \sum_i \left \frac{n_i^k}{n_i} - \frac{n_i^l}{n_i} \right $	Inter-variable
구조유사 척도	$C_{i,j} = \frac{1}{2} \sum_k \left \frac{n_i^k}{n_i} - \frac{n_j^k}{n_j} \right $	단위지역별로 값이 할당됨, 지도화 가능, Inter-region
구조변동 지수	$D_i = \frac{1}{2} \sum_k \left \left(\frac{n_i^k}{n_i} \right)^{l2} - \left(\frac{n_i^k}{n_i} \right)^{l1} \right $	단위지역별로 값이 할당됨, 지도화 가능, 서로 다른 2개의 시점 비교
입지변동 지수	$D_k = \frac{1}{2} \sum_i \left \left(\frac{n_i^k}{n_i} \right)^{l2} - \left(\frac{n_i^k}{n_i} \right)^{l1} \right $	속성별로 단일 값을 가짐, 서로 다른 2개의 시점 비교

역 i 와 j 가 갖고 있는 구조적 유사성을 평가하는 계수이다.

구조변동 지수(Measure of Structural Change, Coefficient of Regional Redistribution)는 특정 두 시점 t_1, t_2 사이 지역 내 산업 구조의 변동을 살펴보는 방법이고, 입지 변동 지수(Index of Locational Change, Coefficient of Industrial Change)는 일정 기간동안 특정 산업의 공간적인 분포가 어느 정도 바뀌어졌는가를 측정하는 척도이다.

2.3. 지역격차 분석

지역격차(Regional Inequality)란 일반적으로 지역발전 수준의 차이를 일컫는 개념

으로서, 어떤 한 시점을 기준으로 단위지역 간에 존재하는 사회경제적 발전의 분산 혹은 변이의 정도라고 정의되어질 수 있다(김안제, 1974). 지역격차 분석은 인구, 산업, 소득, 보건, 복지, 교육, 환경 등 분석 대상 변수의 선택에서 출발한다(Cowell, 2000). 지역격차 분석에 관한 연구들은 관심 부문에 해당하는 특정 변수를 추출하거나, 여러 부문 변수들을 합성하여 지역격차 지수를 산출하는 것이 일반적이다. 지역격차의 정량화는 [표 2]와 같은 수학적 모델에 근거한다.

지역격차 측정 모델은 그 원리에 따라 분산모델, 조합 모델, 엔트로피 모델, 사회복지함수 모델, 형평성 모델 등으로 분

<표 2> 지역격차 분석 모델

지 표	공 식	비 고
Gini	$G = \frac{1}{2n^2 \bar{y}} \sum_{i=1}^n \sum_{j=1}^n y_i - y_j $	y_i, y_j 는 각 단위지역이 가지는 값. \bar{y} 는 평균, n 은 단위지역의 수
Gastwirth	$\frac{1}{n^2} \sum_{i=1}^n \sum_{j=1}^n \frac{ y_i - y_j }{y_i + y_j}$	y_i, y_j 는 각 단위지역이 가지는 값, n 은 단위지역의 수
Theil	$T = -\sum_{i=1}^N \frac{y_i}{Y} \log\left(\frac{y_i}{Y}\right)$	y_i 는 각 단위지역이 가지는 값. Y 는 전체지역의 총합, N 은 단위지역의 수
Lieberson	$D = 1 - \left(\frac{1 - \sum_{k=1}^n Y_k^2}{1 - \frac{1}{n}} \right)$	Y_k 는 각 단위지역이 가지는 값의 비율, n 은 단위지역의 수
Coulter	$I = \frac{\sqrt{\sum \left(\frac{X_i}{S} - E_i\right)^2}}{\sqrt{1 + (\sum E_i^2) - 2 \min\{E_i\}}}$	S 는 전체 지역의 총합, X_i 는 각 단위지역이 가지는 값, E_i 는 각 단위지역의 기대값
Nagel	$N = \frac{\sum (A - M)^2}{\sum M^2}$	A 는 각 단위지역이 가지는 실제값, M 은 각 단위지역의 최소요구치
Atkinson	$A_\varepsilon = 1 - \left[\frac{1}{n} \sum_{i=1}^n \left(\frac{y_i}{\bar{y}}\right)^{1-\varepsilon} \right]^{\frac{1}{1-\varepsilon}}$	n 은 단위지역의 수, y_i 는 각 단위지역이 가지는 값, \bar{y} 는 평균, ε 는 가중치

류되는데, Gini와 Gastwirth의 모델은 단위 지역이 가진 값들의 편차에 근거하므로 분산모델에 해당하고, Lieberson의 모델은 단위지역이 가진 값들의 조합에 따른 확률에 근거하므로, 조합모델에 해당하며, Theil의 모델은 엔트로피 이론에 근거하므로 엔트로피 모델에 해당하고, Atkinson의 모델은 분배불평등을 측정하기 위해 고안된 것이므로 사회복지함수 모델에 해당하며, Nagel과 Coulter의 모델은 필요량과 실제량의 차이에 근거하므로 형평성 모델에 해당한다(박성준, 1996).

2.4. 상관관계 분석

행정구역 단위의 사회경제적 자료를 분석함에 있어서, 두 변수 사이의 상관관계를 파악하는 방법으로 널리 채택되는 것은 Pearson의 상관계수와 Spearman의 순위 상관계수이다.

2.5. 공간 분산분석

공간 분산분석(Spatial ANOVA)은 전체 지역을 몇 개의 영역으로 분할하였을 때 각 영역간에 특정 변수 값들의 분포가 유의한 차이를 나타내는지를 확인하고 검증하는 통계기법이다(Griffith, 1992). 공간 분산분석은 한편 공간자료의 회귀모형의 한 형태로 볼 수 있는데, 이때 각 영역을 지정하는 코드(indicator) 값을 회귀모형의 범주형 독립변수인 지시(dummy) 변수로 이해하면 된다.

각 영역 (예를 들어, 서울시의 강남/강북 영역구분)은 전체지역의 공간적 부분

집합이며, 이의 구분은 범주형 변수를 통해 입력하거나 사용자가 지도상에서 해당 단위 지역들을 클릭하여 대화형으로 선택을 할 수 있다. 통상적인 ANOVA의 용어를 빌린다면 이 영역의 구분은 “처리(treatment)”에 해당하며, 공간 분산분석의 경우 각각의 처리는 공간적인 분할 영역이 된다. 회귀분석의 결과에 포함된 정보 중에 지시변수의 관련 계수는 지역별 평균 값에 대한 차이를 나타내며, 해당 t-value (혹은 p-value)는 지역별 차이에 대한 통계적 유의성을 나타낸다.

2.6. 회귀 분석

상관관계 분석이 두 변수가 어느 정도로 밀접하게 연관되어 있는가의 상관정도를 파악하는 것이라면, 회귀분석은 변수들 간의 상호관련성을 표현하는 관계식을 설정하여 특정변수의 변화에 따라 다른 변수가 어떻게 변화하는지를 파악하는 것이다. 잔차는 회귀방정식에 의한 예측치(\hat{Y})와 실제치(Y)와의 차이를 의미하며, 선정된 x 변수에 의해 설명되어지지 않은 Y 의 분산을 설명하는 데 도움을 준다. 지역이 가진 값에 근거하여 회귀분석을 수행할 경우에는 각 단위지역별로 잔차를 구할 수 있으며, 잔차의 분포도를 통해 Y 변수의 공간적 자기상관이 존재하는지의 여부를 시각적으로 확인할 수도 있다(이희연, 1990).

잔차도(Residual Map)는 통계GIS의 중요한 레이어에 속하며 이를 통해 시각화할 수 있는 잔차의 공간적 분포패턴은 새로운 가설을 세우거나 과대/과소 추정지역의 공통적 현상을 살펴봄으로서 모형의 추가

설명변수를 찾아내는 단서를 제공한다.

2.7. 공간적 자기상관 분석

공간적 자기상관이란 지표면 상의 특정 지역에서 발견되는 활동이나 현상이 주변 지역과 유사한 정도로 연관되어 있는 현상이다. 대부분의 공간적 자기상관의 측도는 지리적 인접도에 관한 행렬과 각 공간 객체의 속성값의 유사도 행렬에 의해 산출된다. Moran과 Geary의 지수들이 보편적으로 계산되며 국지적 규모에서 공간적 상관관계를 분석할 수 있는 LISA(Local Indicator of Spatial Association) 유형의 지수도 최근에 제시되었다.

3. 웹 기반 지역분석 시스템의 설계

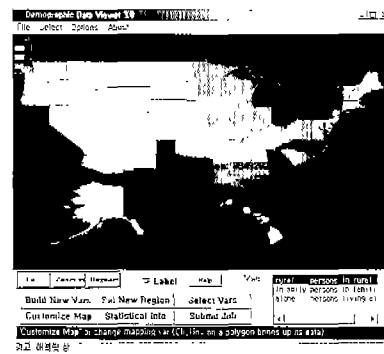
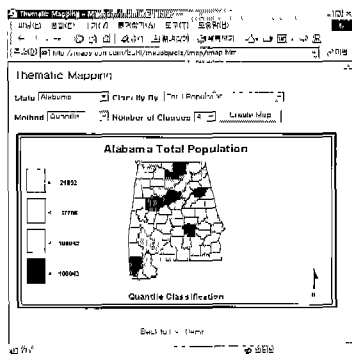
3.1. 시스템의 기능적 요구분석

최근 정부기관이나 지방자치단체에 의

해 공식적으로 발간된 통계자료들은 웹 사이트를 통해 제공되기 시작했다. 그러나 통계청에서 운영하는 Stat-Korea¹⁾를 비롯하여 통계자료를 서비스하는 웹 사이트들은 그 제공방식에 있어서 정적인(static) HTML 형태에 머물고 있으며, 온라인 분석기능은 물론 통계지도가 가능한 사이트는 찾아 볼 수 없었다.

지역단위의 통계자료를 주제도화 (Thematic Mapping)하는 국외 웹 사이트 중에서는, ESRI의 Thematic Map²⁾과 CIESIN(Center for International Science Information Network)의 Demographic Data Viewer³⁾가 비교적 양질의 서비스를 제공하고 있다. 이들은 주 (state) 단위로 인구 센서스 자료의 단계구분도를 제공하며, Demographic Data Viewer의 경우, 한 번의 작업에 2개 이상의 변수를 선택하여 분석할 수 있는 기능도 갖추었다.

기존의 시스템을 벤치마킹한 결과 본 연구에서 목적으로 삼는 지역분석 시스템의 요구사항은 다음과 같다.



[그림 1] Thematic Map(左)과 Demographic Data Viewer(右)

1) <http://www.stat.go.kr>
 2) <http://www.esri.com>
 3) <http://www.ciesin.org>

자료 선택의 유연성: 지역분석에 사용되는 사회경제적 자료는 그 종류가 매우 다양하므로, 모든 범주의 변수는 한번의 작업으로 접근 가능해야 하며, 기존 변수의 조합과 연산에 의해 새로운 변수를 생성하는 기능이 필요하다.
 자료 교환의 유연성: 서버에서 제공되

는 자료를 일방적으로 제공받는 것이 아니라 사용자 자신이 보유한 자료를 서버에 전송하여, 이 데이터를 바탕으로 공간통계 분석을 수행할 수 있도록 하는 양방향의 자료 교환이 가능해야 한다.
 - 가시화 형태의 다양성: 지역분석의 대상이 되는 자료와 지역분석의 결과로서

<표 3> 시스템의 기능적 구성

주기능	세부기능	기능사양
분석대상의 설정	변수의 선택	탐색기 형태의 변수 선택 화면
	변수의 변형과 합성	서버에서 제공되는 데이터 변수들의 조합과 연산을 통해 새로운 변수를 생성
사용자 데이터의 처리	사용자 데이터의 전송	서버에서 제공되는 데이터셋 이외에 사용자 자신이 보유한 자료를 분석하기 위한 사용자 자료의 전송
	사용자 데이터의 구문해석	정의된 포맷을 준수하는 사용자 데이터에 대한 구문 해석(parsing)
	사용자 데이터의 DB저장	구문해석된 사용자 자료를 DB에 저장함으로써 서버에서 제공되는 자료와 동일한 방법으로 이용할 수 있도록 함
지역분석 통계량의 산출	기술적 통계량의 산출	평균, 표준편차 등의 기술적 통계량 산출
	Z-Score의 산출	단위가 상이한 변수들을 비교 가능한 척도로 변환
	변수합성을 위한 가중치 설정	2가지 이상의 변수를 합성하여 단일한 척도로 만들기 위해 각 변수에 대한 가중치를 부여
	지역구조 지수의 산출	입지계수, 집중계수 등 지리적 사상의 집중/분산 정도를 측정하는 지수를 산출
	지역격차 지수의 산출	변동계수, Gini 계수 등 지역격차의 정도를 요약하는 지수를 산출
	상관관계 분석	변수들의 조합에 의한 두 변수 간의 상관관계를 분석 (Pearson, Spearman)
	회귀분석과 공간 분산분석	선형모형에 의한 단순회귀분석 및 잔차도
	공간적 자기상관 분석	①공간적 분포의 자기상관성에 대한 분석 ②국지적 규모의 공간적 자기상관 분석
자료의 가시화	표 형태의 가시화	분석대상 자료와 분석결과 자료를 표 형태로 list-up
	도표 형태의 가시화	분석대상 자료와 분석결과 자료를 도표 형태로 디스플레이
	엑셀 형태의 가시화	분석대상 자료와 분석결과 자료를 엑셀 형태로 디스플레이
	지도 형태의 가시화	분석대상 자료와 분석결과 자료를 단계구분도 형태로 디스플레이

산출된 자료는 표, 도표, 엑셀, 지도 등 다양한 형태로 가시화될 필요가 있다.

3.2. 시스템 프로토타입

(1) 시스템의 기능적 구성

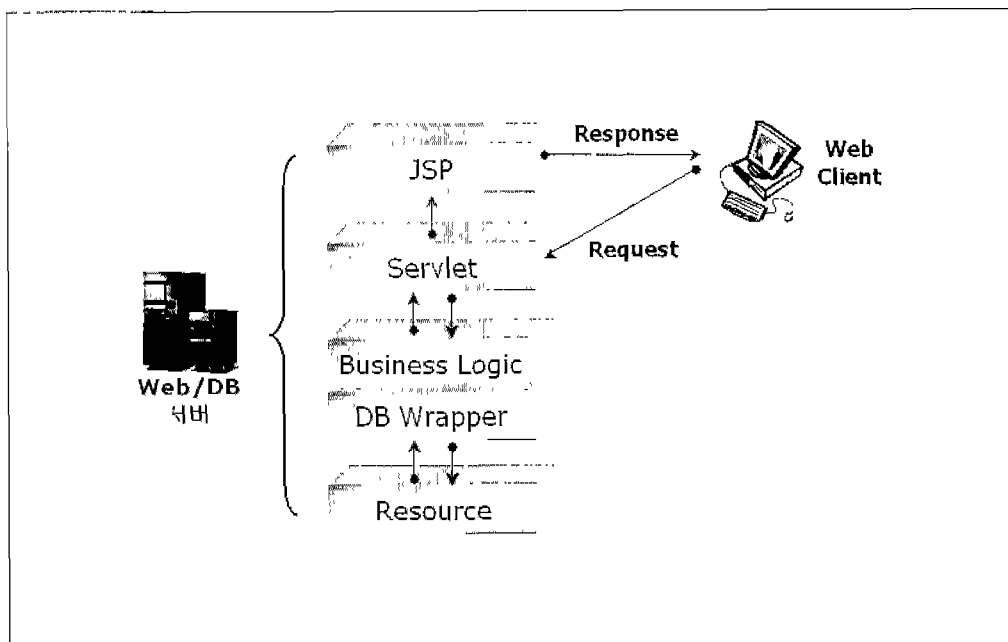
시스템의 기능적 요구분석을 기초로 하여 도출된 이 시스템의 기능사양은 [표 3]과 같이 정리될 수 있다.

(2) 시스템 아키텍처

이 시스템은 웹과 데이터베이스의 연동

에 기초한다. 웹 브라우저를 통해 사용자의 request가 서블릿(servlet)에 전달되면, 서블릿은 Business Logic과 DB Wrapper를 통해 데이터베이스 자원에 접근하고, 참조된 데이터의 분석결과는 JSP를 통해 사용자의 웹 브라우저로 전달된다. 이러한 일련의 과정은 WAS(Web Application Server)인 WebLogic Server가 중개(Brokerage)한다.

[그림 2]에서 보듯이, 이 시스템은 ① 사용자의 Request를 분석하는 부분, ② 통계분석의 Business Logic을 처리하는 부분, ③ DB의 제어를 담당하는 부분, 그리고 ④ 사용자에게 Response를 내보내는 부분을 캡슐화하여 각 전담 기능의 객체지향성과 성능 향상을 추구한다. 이 시스템의 구성을 위한 소프트웨어 환경은 [표 4]와 같다.

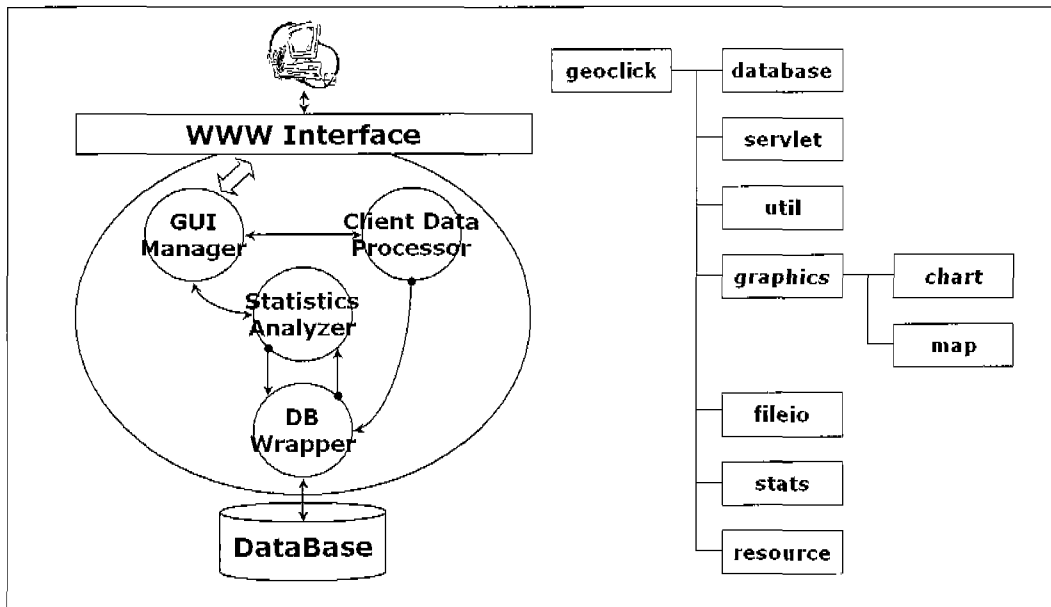


[그림 2] 시스템 아키텍처

<표 4> 시스템의 소프트웨어 환경

역 할	사용된 소프트웨어
DBMS	Oracle 8i ver. 8.1.6 Enterprise Edition
WAS	WebLogic Server ver. 5.1.0 Base Edition
Development Tool	JDK ver. 1.2.2_007 Standard Edition

고, 통계분석의 환경설정과 result set에 대한 디스플레이 기능을 담당한다. Statistics Analyzer는 선택된 변수에 해당하는 데이터가 fetch되면, 사용자가 지정한 통계분석 방법에 기초하여 데이터를 연산, 가공함으로써 통계량을 산출한다. DB Wrapper의 역할은 DB 접속과 SQL 처리, result set의 생성에 있다. 그리고 Client Data Processor는 원격지 사용자의 자료를 전송



[그림 3] 시스템 핵심 모듈(左)과 geoclick package의 구성도(右)

3.3. 시스템 핵심 모듈의 구성

웹에 기반한 지역분석의 핵심 모듈은 [그림 3]과 같이 GUI Manager, Statistics Analyzer, DB Wrapper, Client Data Processor 등으로 구성된다. GUI Manager는 분석 대상 데이터셋에 대한 탐색과 list-up, 그리

받아 이를 구문해석하여 DB에 저장함으로써, 사용자 자료가 서버에서 제공하는 데이터셋과 동일한 형태로 이용 가능하도록 처리하는 역할을 수행한다. 이러한 핵심 모듈을 비롯하여 모든 Java 클래스들은 geoclick이라는 이름으로 패키지화 하였다.

(1) 사용자 인터페이스 제어 모듈

사용자 인터페이스 제어 모듈은 역할은 크게 두 가지로 구분된다. 첫 번째는 사용자가 선택한 데이터셋과 지정 조건을 request 객체 형태로 변환하여 통계분석 제어 모듈에 전달하는 것이고, 두 번째는 통계분석 결과를 표, 도표, 엑셀, 지도의 형태로 웹 브라우저 상에 디스플레이하는 것이다.

(2) 사용자 데이터 제어 모듈

사용자는 서버에서 제공하는 데이터 이외에 자신이 보유한 데이터를 이용하여 지역분석을 수행하고자 하는 요구를 가질 수 있다. 이러한 요구에 대응하여, 사용자 데이터 제어 모듈은 사용자가 업로드한 데이터 (ASCII 파일)를 구문해석하고, 이를 데이터베이스에 저장할 수 있는 객체로 변환하여 데이터베이스 제어 모듈에 전달한다.

(3) 통계분석 모듈

통계분석 모듈은 사용자 인터페이스 제어 모듈로부터 수신된 사용자 요청사항을 참조하여 변수 리스트를 데이터베이스 제어 모듈에 전달하고, 데이터베이스 제어 모듈의 SQL 처리 결과를 result set 객체로 전달 받아, 통계분석을 위한 연산을 수행한다. 연산 결과로 산출된 통계량은 사용자 인터페이스 제어 모듈을 통해 디스플레이된다.

(4) 데이터베이스 제어 모듈

데이터베이스 제어 모듈의 역할은 크게 두 가지로 구분된다. 첫 번째는 사용자 데이터 제어 모듈로부터 전달받은 사용자 자료를 데이터베이스에 저장할 수 있도록 SQL문을 구성하고 실행하는 것이다. 두 번째는 통계분석 모듈로부터 전달된 변수 리스트를 SQL문으로 변환하여 데이터베이스에 전달하고, SQL문 수행 결과를 다시 통계분석 모듈에 반환하는 것이다. 이러한 두 가지 기능은 동적인 SQL문 생성을 통해 이루어진다.

4. 웹 기반 지역분석 시스템의 구현

4.1. 데이터베이스의 구축

본 연구에서는 행정구역 단위의 사회경제적 자료를 이용한 지역분석을 수행하기 위해 서울시 행정 구역 단위 자료의 데이터베이스를 구축하였다. 자료원은 1995년 인구 및 주택 총조사 원시자료(약 1,400만 레코드)와 1995년 사업체기초통계 원시자료(약 70만 레코드)를 사용하였다.

방대한 양의 원시자료를 데이터베이스를 통해 집계하는 작업은 ① 데이터 로딩 (loading), ② 트리거 구성, ③ 내장 프로시저 (stored procedure) 실행과 같은 절차를 거친다. 데이터 로딩 단계에서 원시자료는 Oracle SQL*Loader를 통해 데이터베이스 테이블에 저장된다. 각 원시자료는 1인, 1가구, 1사업체가 각각의 레코드로 구성된 미시자료이므로, 이를 구 단위별로

집계하는 전처리 과정은 트리거와 내장 프로시저 기법을 필요로 한다. 트리거 기법은 레코드의 insert 이벤트에 반응하는 Oracle 객체를 구성하여 집계 테이블에 지역별/항목별로 수치의 누적이 이루어지도록 하는 것이고, 내장 프로시저 기법은 모든 레코드의 insert가 완료된 후에, 지역별/항목별 수치의 누적이 이루어지도록 하는 PL/SQL을 작성하여 이를 명시적으로 실행하는 것이다.

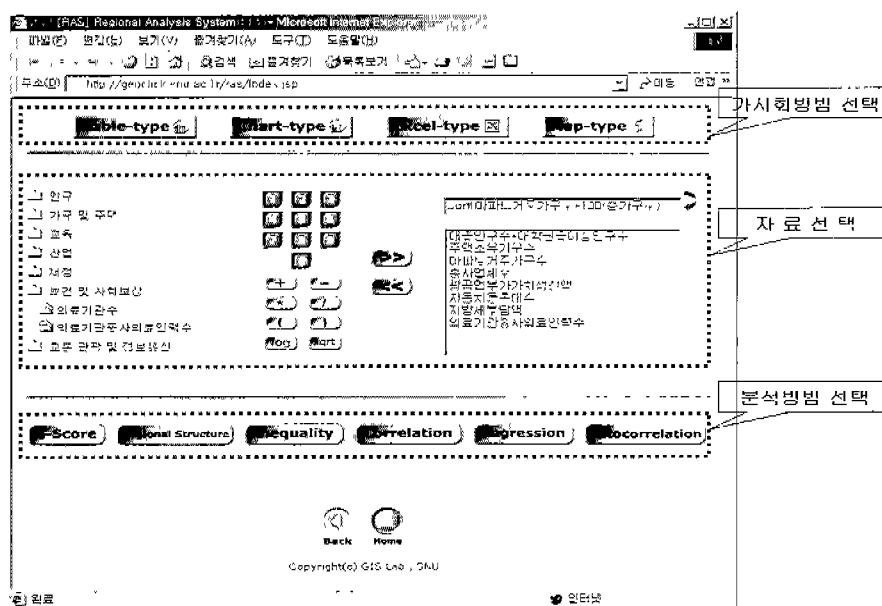
4.2. 사용자 인터페이스의 구현

이 시스템의 기본적인 사용자 인터페이스는 [그림 4]와 같이, 자료 선택 부분, 분석방법 선택 부분, 가시화방법 선택 부분으로 구성된다. 자료 선택 부분에서는 예를 들어, $\sqrt{\text{아파트거주가구수} \times 100 / \text{총가구수}}$ 과 같이 기존 변수의 조합과 연산

에 의해 새로운 변수를 생성하는 기능이 지원된다. 사용자가 지정한 변수의 조합과 연산은 구문해석기에 의해 Oracle의 sum, count, avg, sqrt 등의 함수로 대체되어 SQL문에 포함될 column alias로 변환된다.

(1) 표 형태의 가시화

이 시스템은 기본적으로, 사용자의 request를 처리하는 서블릿과 분석결과 response를 처리하는 JSP의 연계구조로 구성된다. 디스플레이할 result set은 GeoclickAttributes Container 클래스를 통해 JSP로 전달되며, JSP는 GeoclickResultSet으로부터 추출한 정보를 HTML 태그 사이에 삽입함으로써 테이블 형태의 웹 문서를 생성한다. geoclick.database 패키지와 geoclick.util 패키지에서 표 형태의 가시화와 관련된 클래스는 다음과 같다.



[그림 4] RAS(Regional Analysis System)의 기본 화면

주요 클래스	기 능
GeoclickResultSet	java.sql.ResultSet을 implements하며, SQL 실행 결과를 보관
GeoclickProperties	java.util.Properties를 extends하며, JSP로 전달할 정보를 보관
GeoclickAttributes	GeoclickResultSet과 GeoclickProperties를 JSP로 전달하기 위한 저장소
GeoclickAttributes Container	GeoclickAttribute를 보관할 Vector 형태의 객체로서 setURL 메소드를 통해 포워딩(forwarding)할 JSP의 주소를 설정

(2) 도표 형태의 가시화

geoclick.graphics.chart 패키지 클래스와 geoclick.graphics.encoder 패키지 클래스들은 도형 객체를 그리고(draw), GIF-Encoding을 수행함으로써 GeoclickResultSet의 정보를 도표로 변환한다.

주요 클래스	기 능
GeoclickChartAxis	표현항목의 수, 레코드의 수, 레코드의 최대최소값 등을 참조하여 도표의 스케일을 조정
GeoclickChartColors	표현항목의 수에 따라 각 항목별 색상 값을 부여
GeoclickChartLegend	표현항목의 수에 따라 범례를 설정
GeoclickChart	도표의 스케일과 레코드 값에 따라 bar, line, circle 형태의 도표를 생성
GeoclickEncoderGIF	GeoclickChart 객체를 GIF 포맷으로 encoding

(3) 엑셀 형태의 가시화

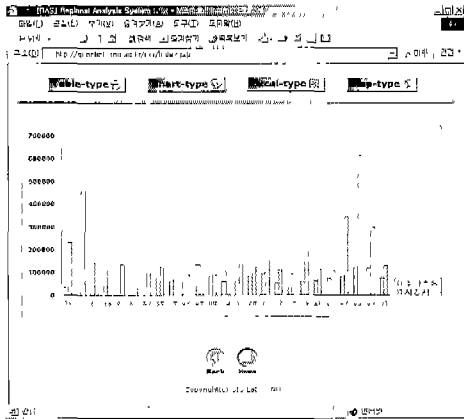
GeoclickResultSet에 보관된 정보는 CSV (comma 구분자를 가진 텍스트 파일)와 같은 엑셀 호환 파일로 기록될 수 있다. geoclick.fileio 패키지의 GeoclickCSVWriter 클래스는 java.io.BufferedWriter 클래스와 java.io.BufferedWriter 클래스를 이용하여 GeoclickResultSet에 보관된 정보를 CSV 파일로 변환한다.

(4) 지도 형태의 가시화

geoclick.graphics.map 패키지는 점, 선, 면 등의 도형 객체를 생성하고, GeoclickResultSet에 보관된 정보를 추출하여 각 지역에 해당하는 도형 객체에 색상값을 부여함으로써 단계구분도를 작성한다.

주요 클래스	기 능
GeoclickPoint	Point 도형객체
GeoclickPointSet	Point set 도형객체
GeoclickLine	Line 도형객체
GeoclickPolyLine	Polyline 도형객체
GeoclickPolygon	Polygon 도형객체
GeoclickLegendEditor	단계구분도에 부여할 color spectrum을 정의
GeoclickGroupingManager	equal interval, standard deviation 등의 급간 구분 방식을 관리

표, 도표, 엑셀, 지도 형태의 가시화 예는 [그림 5]와 같다.



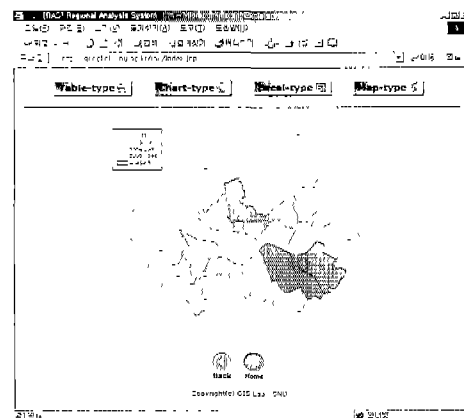
표형태의 가시화

[신원번호 List up]

번호	성명	주민번호	생년월일	성별	직업	주소	전화번호	이메일
11010	김민	910101	1970.01.01	남	학생	서울시 강남구	02-1234-5678	kimmin@naver.com
11020	이영	920202	1975.02.02	여	회사원	서울시 중구	02-2345-6789	leeyoung@naver.com
11030	박철	930303	1980.03.03	남	무직	서울시 동구	02-3456-7890	parkcheol@naver.com
11040	정혜	940404	1985.04.04	여	주부	서울시 서구	02-4567-8901	jeonghye@naver.com
11050	최준	950505	1990.05.05	남	학생	서울시 남구	02-5678-9012	choijun@naver.com

도표 형태의 가시화

엑셀 형태의 가시화



지도 형태의 가시화

[그림 5] 분석대상과 분석결과의 가시화 유형

4.3. 데이터 처리기의 구현

(1) 질의 생성기

지역분석 시스템에서 제공하는 행정구역 단위의 사회경제적 자료는 다양한 범주로 구성되므로, 다중 테이블에 분산된 형태로 저장된다. 이러한 데이터셋 중에서 사용자가 어떤 테이블에 있는 어떤 변

수를 선택할 지는 예측할 수 없으므로, 사용자 선택사항을 처리하는 질의 모듈은 분산 테이블의 다중 조인(multiple join)을 동적으로 처리할 수 있는 질의 생성 구조를 갖추어야 한다. 다중 조인은 행정구역코드를 기준으로 이루어지며, geoclick.database 패키지의 클래스들은 사용자가 선택한 변수에 해당하는 테이블 이름과 칼럼 이름의 리스트를 전달받고, 이를 해석, 재배치


```

=== 실행 ===
private String makeColumnList(Vector columns) {
    String columnList = "";
    for(int i = 0; i < columns.size(); i++) {
        columnList += columns.elementAt(i).toString();
        if(i < columns.size() - 1)
            columnList += ",";
    }
    return columnList;
}

private String makeTableList(Vector tables) {
    String tableList = "";
    for(int i = 0; i < tables.size(); i++) {
        tableList += tables.elementAt(i).toString();
        if(i < tables.size() - 1)
            tableList += ",";
    }
    return tableList;
}

private String makeJoin(Vector tables) {
    String joinList = "";
    for(int i = 0; i < tables.size(); i++) {
        joinList += PRIMARY_TABLE + "." + PRIMARY_KEY + "="
            + tables.elementAt(i).toString() + "." + PRIMARY_KEY;
        if(i < tables.size() - 1)
            joinList += "AND ";
    }
    return joinList;
}
=== 유략 ===

```

[그림 6] 질의 생성기 프로그램의 일부

함으로써 동적인 질의문을 구성한다.

(2) 사용자 데이터 구문해석기

주요 클래스	기 능
GeoclickVariableParser	java.util.StringTokenizer 클래스를 이용하여 사용자 지정 변수에 해당하는 테이블명과 칼럼명의 리스트를 생성하고, 이를 GeoclickSqlBuilder로 전달
GeoclickSqlBuilder	GeoclickVariableParser로부터 전달받은 리스트를 이용하여 칼럼 이름에 alias 부여하고, 예약어 (reserved word)를 배치함으로써 SQL 문을 생성

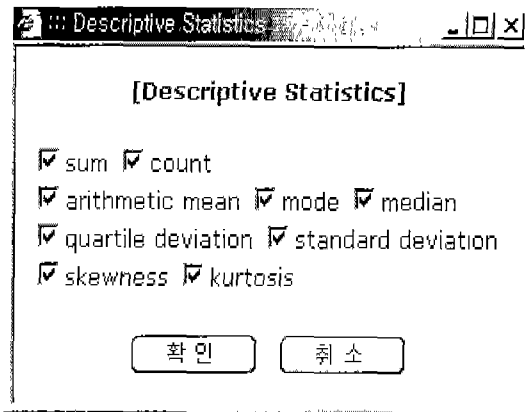
사용자 데이터 구문해석기는 Oracle의 SQL-Loader와 같은 기능 수행하며, 사용자가 전송한 파일을 구문해석하여 데이터베이스에 저장하기 위한 객체의 형태로 변환한다. 이때, 사용자가 전송할 수 있는 파일은 ASCII 텍스트 포맷으로 제한하며, 사용자는 파일 구성에 사용된 구분자를 지정할 수 있다. geoclick.fileio 패키지의 GeoclickASCIIParser 클래스는 사용자가 지정한 구분자를 기준으로 전송 파일을 구문해석하여 데이터베이스 테이블에 저장될 칼럼 수와 레코드 수를 계산하고, value list를 구성하여 GeoclickSqlBuilder로 전달한다. geoclick.database 패키지의 GeoclickSqlBuilder 클래스는 create table

DDL(Data Definition Language)와 insert DML(Data Manipulation Language)를 이용하여 value list를 데이터베이스 테이블에 저장한다.

4.4. 지역분석 통계모듈의 구현

(1) 기술적 통계량

산술평균, 최빈치, 중위수 등의 대표값과, 표준편차, 사분위편차, 그리고 왜곡도, 첨도 등의 기술 통계량은 사용자 지정에 의해 모두 또는 선택적으로 구해질 수 있다.



[그림 7] 기술적 통계량 선택 화면

(2) 표준 정규값

지역분석에서 사용되는 사회경제적 자료는 변수의 유형별로 상이한 측정단위를 가지는데, 표준 정규값(Z-Score)은 성질과 단위가 다른 개별 변수들을 비교 가능한 척도로 변환하는 방법이며, $(X-\mu)/\sigma$ 의 공식에 의해 산출된다.

[그림 8] 표준 정규값의 산출의 실행 예

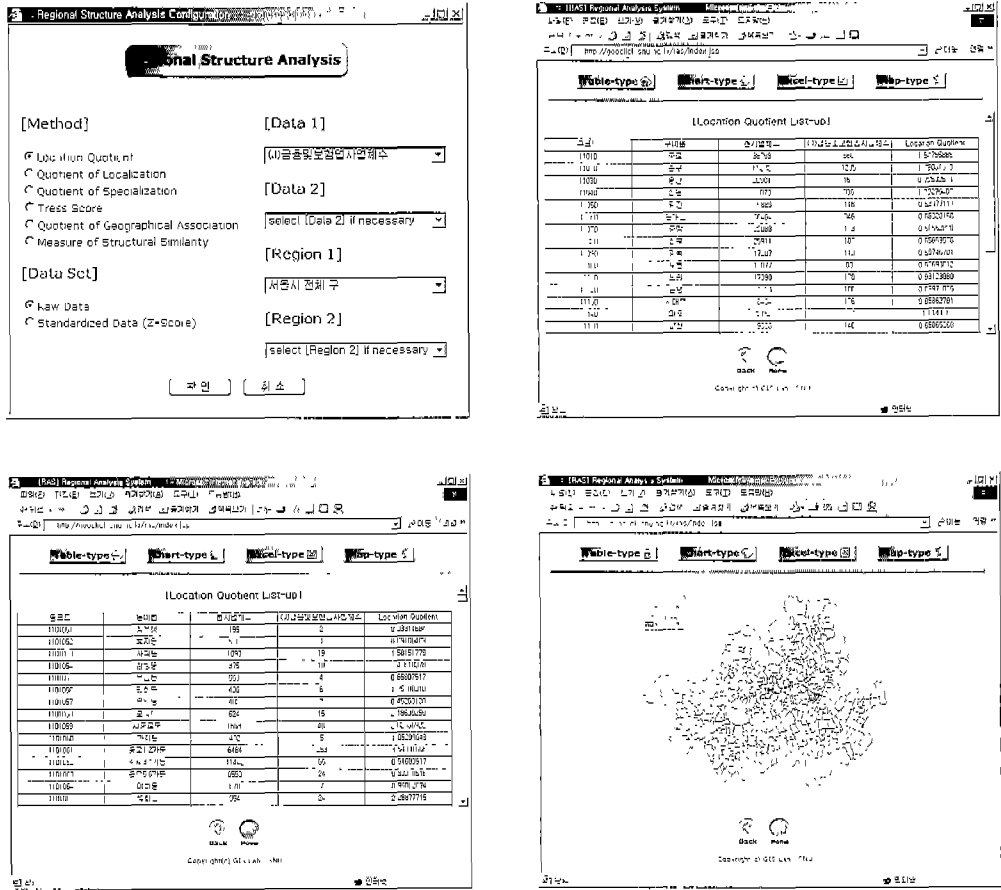
(3) 지역구조 분석

지역구조 분석 방법 중에는 두 종류의 자료를 쌍으로 하여 분석하거나 (예: 입지연관계수), 두 개의 지역을 쌍으로 하여 분석하는 (예: 구조유사척도) 기법이 있으므로, 분석을 위한 환경설정 화면은 필요한 경우에 한하여 자료와 지역을 쌍으로 선택할 수 있도록 구성한다. [그림 9]는 서울시 구별 및 동별 금융 및 보험업 사업체수를 기준으로 하여 입지계수를 산출한 예제이다.

(4) 지역격차 분석

[그림 10]의 실행 예에서 지역격차 측정을 위해 선택된 변수는 (대졸인구수 + 대학원졸업이상인구수), 주택소유가구수, 아파트거주가구수, 총사업체수, 광공업부가치생산액, 자동차등록대수, 지방세부담액, 의료기관종사 의료인력수 등이며, 이

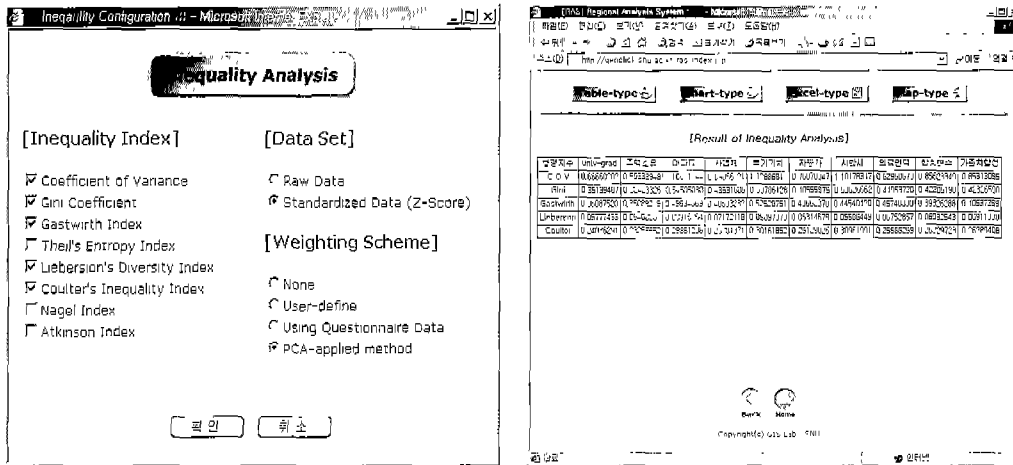
지역분석을 위한 웹 기반 통계GIS 연구



[그림 9] 입지계수의 산출의 실행 예

중 (대출인구수+대학원졸업이상인구수)는 변수의 조합과 연산에 의해 새로이 생성된 변수이다. 이들 변수에 대한 지역격차 측정은 격차지수의 선택, 데이터셋의 선택, 가중치 부여방법의 선택 등 분석을 위한 사용자 옵션을 선택함으로써 이루어

진다. 데이터셋으로 “Standardized Data”를 선택할 경우, 측정단위가 통일되므로 변수의 합성이 가능해지고, 가중치 부여방법 중의 하나를 지정할 경우, 각 변수들에 대한 가중치를 부여한 합성변수에 대한 격차지수가 도출된다.



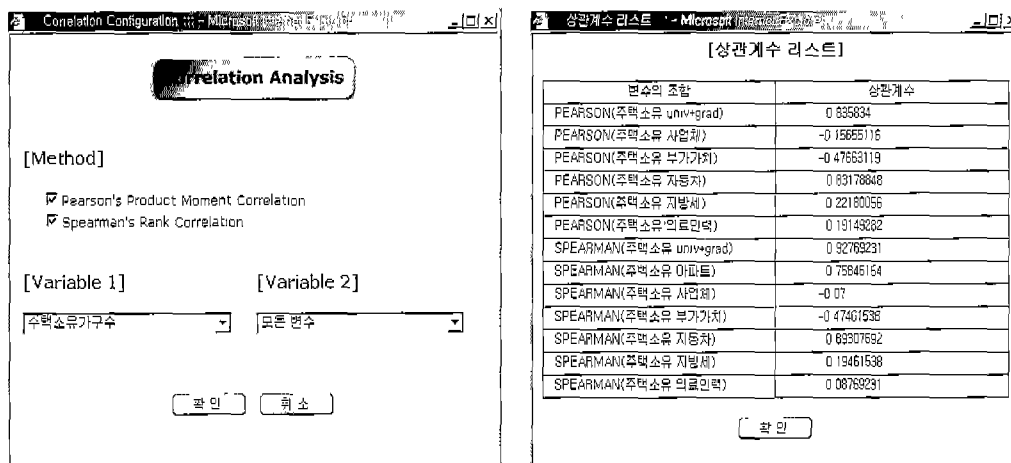
[그림 10] 지역격차 분석의 실행 예

(5) 상관관계 분석

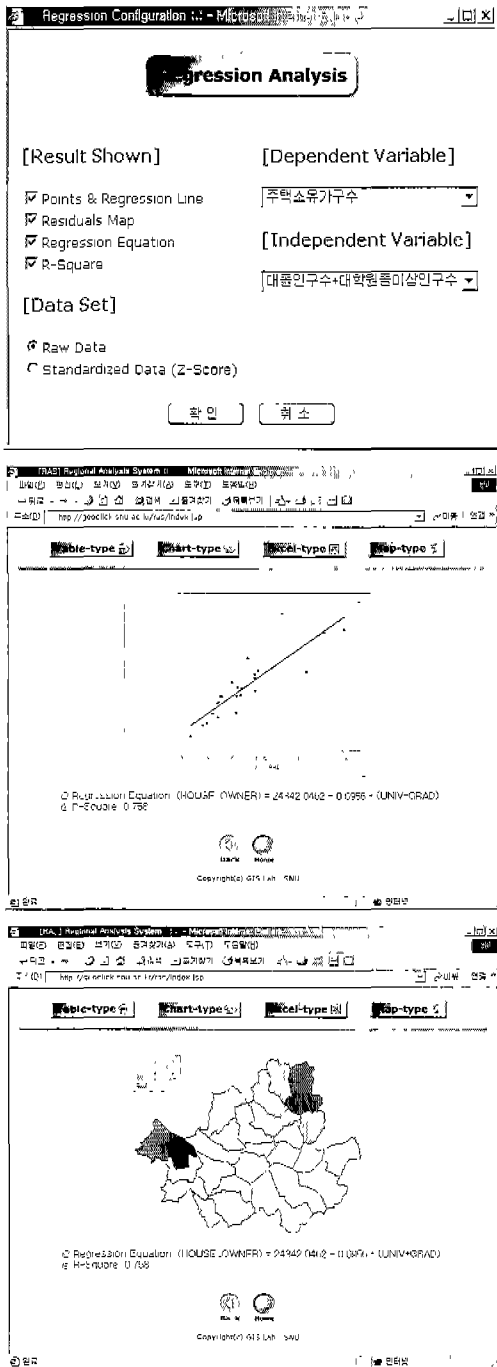
주택소유가구수와 다른 모든 변수들 간의 상관관계를 Pearson과 Spearman의 방법으로 구한 결과는 다음과 같다.

(6) 회귀분석

회귀분석의 결과로서 도출되어지는 것은 회귀선, 잔차도, 회귀방정식, R-Square 등이 있으며, 주택소유가구수에 대한 (대졸인구수+대학원졸이상인구수)의 영향력을 단순회귀모형을 통해 분석한 결과는 다음과 같다.



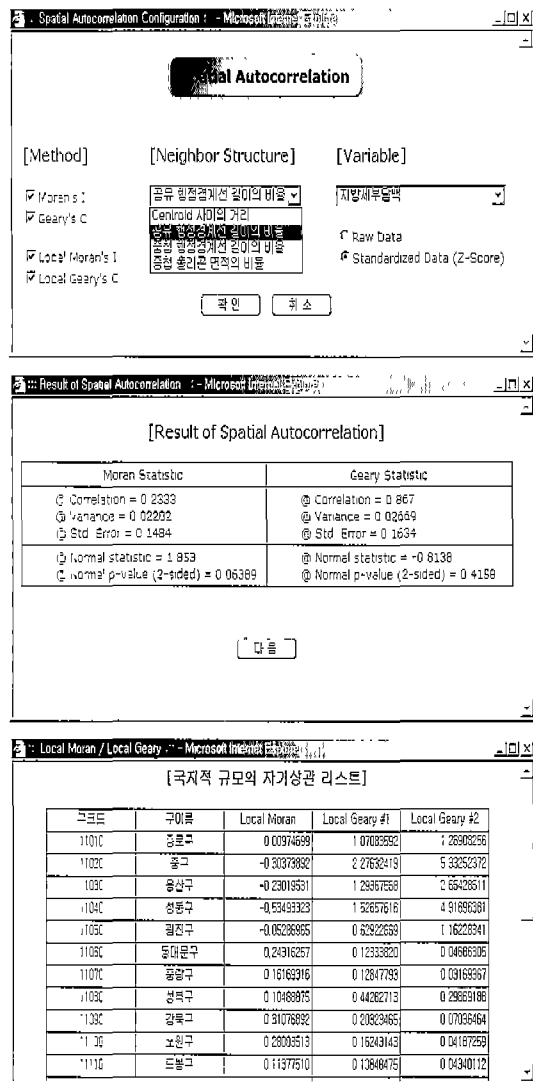
[그림 11] 상관계수 산출의 실행 예



[그림 12] 회귀분석의 환경설정과 분석결과
회귀선 및 잔차도

(7) 공간적 자기상관 분석

공간적 자기상관 분석을 위한 가중치 부여 방식으로는 ①구 중심점 (centroid) 사이의 거리, ②구 경계의 길이 대비 인접 경계의 길이, ③구 경계 길이 대비 중심점 경계 길이, ④구 면적 대비 중첩 폴리곤 면적 등 4가지 선택사항을 제공한다.



[그림 13] 공간적 자기상관 환경설정과 분석결과

5. 요약 및 결론

이 연구에서는 웹 상에서 운영을 기본 전제로 하고 행정구역 단위의 사회경제적 자료와 이의 분석모듈을 동시에 갖추고 있는 통계GIS의 모형을 설계하고 구축하였다. 서울시를 사례지역으로 1,500만 건의 미시자료를 구별/항목별로 효율적으로 집계함으로써 지역분석 시스템의 공간통계를 위한 기초 데이터 풀로 활용하였다. 이 시스템의 지역분석 모듈은 통계GIS의 Business Logic에 해당하는 소프트웨어 컴포넌트로서 이를 통해 지역의 구조 및 격차 분석, 상관관계 분석, 회귀 및 잔차 분석, 공간적 분산분석 등 지역의 인구, 산업, 토지이용, 편익시설의 분포를 분석하는 핵심 통계기법을 데이터와 함께 서비스할 수 있다.

이 시스템은 기존의 웹 지역정보 시스템이 가진 정적인 HTML 구조를 극복하고, 사용자의 request에 동적으로 반응하는 특징을 가질 뿐만 아니라, 공간통계에 대한 기본적인 지식과 간단한 마우스 조작만으로도 온라인 상에서 지역분석을 수행할 수 있는 환경을 제공한다. 웹 환경에서 운영되는 통계GIS의 설계 컨셉트는 자료 선택과 변형의 유연성 과 자료 교환의 유연성, 그리고 가시화 형태의 다양화 등이었으며, 이 요건들은 시범 시스템 구현을 통해 그 효과와 기술적 타당성을 성공적으로 검증할 수 있었다.

이 시스템은 정보기술 측면에서 자바 기반의 JSP와 서블릿의 연결구조로 이루어지며, ①사용자의 request를 분석하는 부분, ②통계분석의 Business Logic을 처리하

는 부분, ③DB의 제어를 담당하는 부분, ④사용자에게 response를 내보내는 부분을 캡슐화함으로써 각 부분의 전담 기능의 객체지향성과 성능 향상을 추구하였다. 특히, DB Wrapper 모듈에서는 예측 불가능한 사용자의 자료선택 조합에 대처하기 위하여 동적인 SQL 생성 기능을 구현하였고, Client Data Processor 모듈에서는 사용자 자료를 구문해석하여 DB화할 수 있는 기능을 구현함으로써 시스템 활용의 폭을 넓히고자 하였다.

본 연구를 통해 제안된 통계GIS이 기존의 유사 시스템과 구별되는 특징점은 미리 저장된 데이터와 분석기능이 사용자와의 양방향 인터페이스를 통해 제한 없이 확장될 수 있다는 점이다. 즉 사용자의 요구에 따라 분석변수의 자유로운 변형과 합성이 연산자 구문해석기를 통해 가능하고, 이를 바탕으로 새로운 지역분석 알고리즘의 구현이 실시간으로 가능하다. 또한 데이터 측면에서는 사용자가 업로드한 자료를 잠정적으로 시스템에 통합하여 확장된 가상 데이터베이스를 만들고 이에 대해 여러 분석을 수행할 수 있다.

웹 기반 통계GIS 모델의 초기평가를 통해 (1) 국가지정통계와 같은 귀중한 자료가 보다 의미있고 실질적인 지역분석 정보로 그 활용도의 제고를 꾀할 수 있고, (2) 주문형 통계지도의 작성이 가능하며, 또한 (3) 정보의 활용 과 더불어 유통의 문제를 동시에 해결할 수 있음이 확인되었다.

통계GIS 프로토타이핑의 성공에 따라 본 시스템의 통계분석기 (Statistics Analyzer)를 CORBA 기반의 IDL로 이식함으로써 웹

환경뿐만 아니라, 기존의 C/S 환경과 새로이 등장하는 모바일 환경에서도 적용 가능한 통계 애플리케이션 서버 (Statistics Application Server)를 개발하는 후속 작업이 진행 중이다.

참 고 문 헌

- 김안제, 1974 지역간 경제적 격차에 관한 축도연구: 한국의 경우를 중심으로, 환경논총, 제1권, 제1호, pp. 27-64.
- 김영표 한선희, 1999, GIS를 이용한 국가통계정보의 활용도 제고방안, 국가지리정보체계 구축 지원연구 25, 국토연구원.
- 박기호, 1997, 웹 기반 지도대수 처리기에 관한 연구, 한국 GIS학회지, 제5권 2호, pp. 147-160.
- 박기호, 2000, 범죄분석 지리정보시스템의 설계와 구현, 한국GIS학회지, 제8권 2호, pp. 213-232.
- 박성복, 1996, 지역발전도 및 지역불평등도의 측정, 한국행정학보, 제31권, 제3호, pp. 165-185.
- 이희연, 1990, 지리통계학: 지리자료분석기법, 법문사.
- 최재선, 1980, 지역경제학, 법문사.
- Chou, Y. 1997 Exploring Spatial Analysis in Geographic Information Systems, OnWord Press.
- Cowell, F. A. 2000 Measuring Inequality, Prentice Hall.
- Darden, J., Tabachneck, A., and Raine, R. 1980 Algorithm 8: Graphic and Mathematical Description of Inequality, Dissimilarity, Segregation or Concentration, Environment and Planning A, 12, pp. 227-234.
- Fischer, M., Scholten, H. and Unwin, D. (eds.) 1996 Spatial Analytical Perspectives on GIS, GISDATA 4, Taylor&Francis.
- Fotheringham, A. and Wegener, M. (eds.) 2000 Spatial Models and GIS, GISDATA 7, Taylor&Francis
- Getis, A. and Ord, K. 1995 Local Spatial Statistics: An Overview in Spatial Analysis: Modeling in a GIS Environment, Cambridge: GeoInformation International.
- Griffith, D. 1992 A Spatially Adjusted N-way ANOVA Model, Regional Science and Urban Economics, 22, pp. 347-369.
- Harder, C. 1998 Serving Maps on the Internet, ESRI, Inc., Redlands:California
- Hall, M. 2000 Core Servlets and Java Server Pages (JSP), Prentice Hall.
- Isard, W. 1960 Methods of Regional Analysis: an Introduction to Regional Science, MIT Press.
- Krueckeberg, D. and Silvers, A. 1974 Urban Planning Analysis: Methods and Models, John Wiley & Sons.
- Lee, J. and Wong, D. 2001 Statistical Analysis with ArcView GIS, John Wiley & Sons.
- Massey, D. and Denton, N. 1988 The Dimensions of Residential Segregation, Social Forces, 67:2, December. pp. 281-315.
- Plewe, B. 1997 GIS ONLINE: Information Retrieval, Mapping, and the Internet, OnWord Press.