인터넷 경제 측정기법 연구

2018. 12. 15

연구수행기관 : 정보통신정책연구원

통계청
Statistics Korea

인터넷 경제 측정기법 연구

2018. 12. 15

연구수행기관 : 정보통신정책연구원

통계청
Statistics Korea

제 출 문

통계청장 귀하

본 연구결과보고서를 **“인터넷 경제 측정기법 연구”**의 최종연구보고서로 제출합니다.

2018년 12월 15일

연 구 기 관 : 정보통신정책연구원
총괄책임자 : 정 용 찬
참여연구원 : 정 현 준
정 부 연
신 우 철
손 넝 선
김 성 환(아주대)

목 차

요약문 ····· 7

제 1 장 서 론 ····· 25
제 1 절 연구의 목적 및 필요성 ····· 25
1. 연구의 목적 ····· 25
2. 연구의 필요성 ····· 25
제 2 절 연구 범위와 연구 방법 ····· 29
1. 연구 범위 ····· 29
2. 연구 방법 ····· 31
3. 연구 추진 경과 ····· 32

제 2 장 인터넷 경제 관련 기존문헌 및 해외사례 연구 ····· 34
제 1 절 인터넷 경제 개념과 범위에 대한 문헌 연구 ····· 34
1. 인터넷(또는 디지털) 경제의 개념 ····· 34
2. 인터넷(또는 디지털) 경제 시장 구모 ····· 37
제 2 절 인터넷 경제 측정 방법론에 대한 사례 연구 ····· 41
1. 인터넷 경제 측정 방법론 개요 ····· 41
2. 인터넷 경제 측정 방법론 관련 주요 사례 분석 ····· 43
3. 인터넷 경제와 플랫폼 경제 간의 연관성 ····· 48

제 3 장 인터넷 경제 측정 방법론 연구 ····· 50
제 1 절 네덜란드 통계청의 인터넷 경제 측정 방법론 적용 연구 ····· 50
1. 네덜란드 통계청(BOD)의 인터넷 경제 측정 방법론 연구 ····· 50
제 2 절 우리나라 데이터 적용 분석 ····· 55
1. 데이터 ····· 55

2. 데이터셋 매칭 전략 ····· 57
3. 데이터 프로바이더 데이터 특성 ····· 60
4. 우리나라 데이터 적용을 위한 수정된 매칭 전략 ····· 61
5. 카테고리 D와 E의 키워드 분석 ····· 65
6. 인터넷 사용에 따른 업체 분류 ····· 72
제 3 절 우리나라 인터넷 경제 측정 결과 ····· 73
1. 데이터 단계별 매칭 결과 ····· 73
2. 연계 데이터 출처별 매칭 결과 ····· 75
3. 인터넷 단계별 매칭 결과 ····· 77
4. 지역별, 산업별 분포 결과 ····· 78
5. 인터넷 단계별 주요 변수 분포 결과 ····· 80
제 4 절 네덜란드 방법론에 대한 한계점 ····· 85
1. 연구의 한계점 ····· 85
2. 인터넷 경제/플랫폼 경제 규모에 대한 측정방법론 검토 ····· 86

제 4 장 플랫폼 기업(비즈니스) 경제 규모에 대한 측정방법론 연구 ····· 88
제 1 절 연구의 필요성 ····· 88
제 2 절 디지털 경제, 인터넷 경제, 플랫폼 경제의 개념과 특징 ····· 88
제 3 절 플랫폼 경제 (양면시장) 이론 ····· 90
1. 양면시장 관련 개념과 사례 ····· 90
2. 양면시장(플랫폼 경쟁) 이론 ····· 91
3. 양면시장 이론의 경쟁정책적 시사점 ····· 93
4. 플랫폼 경쟁의 유형에 따른 시장구조 특성 ····· 95

제 5 장 결 론 ····· 100
1. 주요 연구 내용 및 결과 ····· 100
2. 연구의 한계 및 시사점 ····· 101

참고문헌 ····· 105

〔부 록〕 데이터 프로바이더 제공 데이터의 변수 설명 ····· 108

표 목 차

〈표 1－1〉 빅데이터 국가통계 활용 해외사례 ····· 27
〈표 1－2〉 빅데이터 국가통계 활용 분야 확대 사례 ····· 28
〈표 1－3〉 인터넷 경제 측정 방법론 정립 기본 방향 ····· 31
〈표 1－4〉 인터넷 경제 측정 연구 수행 전략 ····· 32
〈표 1－5〉 연구 추진 경과 ····· 32
〈표 2－1〉 Boston Consulting Group(2010)의 인터넷 경제 범위 ····· 36
〈표 2－2〉 McKinsey(2011)의 인터넷 경제의 범위 ····· 37
〈표 2－3〉 인터넷 경제, 디지털 경제, 플랫폼 경제의 개념 비교 ····· 37
〈표 2－4〉 OECD 주요국 기업들의 웹사이트 or 홈페이지 사용 현황
(2010, 2016) ····· 40
〈표 2－5〉 산출 방식에 따른 인터넷 경제 규모 측정 방법론 ····· 42
〈표 2－6〉 지출 분류별 세부지출항목 ····· 45
〈표 2－7〉 인터넷 관련 지출의 항목 별 활용데이터 ····· 46
〈표 2－8〉 OECD(2012)의 인터넷 경제의 범위 및 추정 방법 ····· 48
〈표 3－1〉 인터넷 경제 측정을 위한 데이터 ····· 52
〈표 3－2〉 네덜란드 사례의 주요 분석 결과 ····· 55
〈표 3－3〉 주요 분석 결과 ····· 56
〈표 3－4〉 매칭 방법론 분류 ····· 58
〈표 3－5〉 세부 유형별 매칭 주요 결과 ····· 61
〈표 3－6〉 통신망식별번호 및 지역번호 할당 현황 ····· 64
〈표 3－7〉 주요 서비스별 번호체계 ····· 65
〈표 3－8〉 네덜란드 인터넷 경제 분류의 카테고리 D와 E ····· 66
〈표 3－9〉 D&E에 속하는 키워드 종류 ····· 69
〈표 3－10〉 국내 인터넷 경제 분류의 D와 E에 대한 표준산업분류 ····· 70

〈표 3－11〉 데이터 매칭 단계별 매출액 및 종사자 수 ····· 75
〈표 3－12〉 연계 데이터 출처별 기업체 및 사업체 수 ····· 76
〈표 3－13〉 연계 데이터 출처별 매출액 및 종사자 수 ····· 77
〈표 3－14〉 인터넷 단계별 기업체 수, 매출액 및 종사자 수 ····· 78
〈표 3－15〉 전체기업 및 인터넷기업 지역별 분포 비교 ····· 78
〈표 3－16〉 전체기업 및 인터넷기업 산업별 분포 비교 ····· 79
〈표 3－17〉 인터넷 단계별 종사자 규모별 기업 분포 ····· 81
〈표 3－18〉 인터넷 단계별 업력별 기업 분포 ····· 82
〈표 3－19〉 인터넷 단계별 지역별 기업 분포 ····· 82
〈표 3－20〉 인터넷 단계별 산업별 기업 분포 ····· 84
〈표 5－1〉 인터넷경제 카테고리 분류별 판정 난이도 ····· 103

그 림 목 차

〔그림 1－1〕 제2차 국가통계발전 기본 계획 추진방향 ····· 29
〔그림 1－2〕 과제 수행체계 ····· 30
〔그림 2－1〕 OECD(2012)의 인터넷 경제의 파급효과의 범위 규정 ····· 35
〔그림 2－2〕 McKinsey(2011)의 인터넷 경제의 규모 ····· 38
〔그림 2－3〕 BCG(2012)의 인터넷 경제의 규모 ····· 38
〔그림 2－4〕 OECD 주요국의 ICT 부가가치 비중(2015) ····· 39
〔그림 2－5〕 OECD 주요국의 인터넷 사용률(2016) ····· 40
〔그림 2－6〕 총지출 기반 접근방식으로 추정된 호주의 GDP 대비 부문별 인터넷 경제 비중 ····· 44
〔그림 2－7〕 총소득 기반 접근방식으로 추정된 호주 인터넷 경제의 업종별 종사자 수 및 부가가치 총액 ····· 44
〔그림 2－8〕 인터넷 경제규모 측정의 세 가지 접근 방식 ····· 47
〔그림 3－1〕 인터넷 활용에 따른 기업 분류 ····· 51
〔그림 3－2〕 데이터 연계 방법론 ····· 53
〔그림 3－3〕 웹사이트와 사업체 기업 연계 방법 ····· 53
〔그림 3－4〕 웹사이트 정제 후 최종 기업체 수 ····· 54
〔그림 3－5〕 한국 데이터를 이용한 웹사이트와 사업자등록번호, 법인등록번호 간 관계 ····· 60
〔그림 3－6〕 우리나라 데이터 적용 분석을 위한 매칭 전략 ····· 62
〔그림 3－7〕 우리나라 번호체계 ····· 63
〔그림 3－8〕 인터넷 사용에 따른 업체 분류 ····· 73
〔그림 3－9〕 데이터 매칭 단계별 기업체 및 사업체 수 ····· 74

요 약 문

1. 연구명

인터넷 경제 측정기법 연구

2. 연구기간

2018. 6. 14~2018. 12. 15

3. 연구의 목적 및 필요성

가. 연구의 목적

○ 인터넷 경제(디지털 경제) 측정에 대한 기존 연구와 해외사례를 분석하고 네덜란드 통계청(CBS)의 인터넷 경제 측정 방법론을 우리나라 통계청 데이터를 적용하여 시산하며 네덜란드 방법론에 대한 한계점 및 대안을 검토하여 우리나라 상황에 맞는 인터넷경제 측정방법을 제시

나. 연구의 필요성

○ 최근 인터넷, ICT 등의 발달로 경제 환경이 산업경제에서 인터넷 경제 또는 디지털 경제로 빠르게 전환하는 가운데 디지털 경제의 비중이 갈수록 높아지는 추세

○ 인터넷은 기본적으로 컴퓨터 네크워크로서 모든 컴퓨터를 하나의 통신망으로 연결해 주는 글로벌 네크워크를 의미하며, 인터넷 이용자수, 인터넷 속도 등 인터넷 기술의 단순한 활용을 측정하는 것과 인터넷 경제를 측정하는 것은 차이가 있음

○ 인터넷의 확산으로 경제전반에 미치는 영향력이 확대됨에 따라 향후 적절한 정책수립을 위해 인터넷경제의 규모와 경제적 파급효과를 추정하기 위한 방

법론 연구의 필요성이 확대되고 있음

○ 인터넷 경제를 계량적으로 측정할 경우 인터넷이라는 통신수단과 관련된 모든 산업 활동과 인터넷에 기반해 수행되는 모든 경제활동을 포함할 것인지 아니면 인터넷을 통해 거래되는 재화나 서비스 또는 인터넷을 통한 금융적 거래 활동만을 포함할 것인지에 대한 개념 정립이 매우 중요

○ 인터넷 경제의 중요성에도 불구하고 그 규모를 측정한 사례가 드물고, GDP에서도 독립된 분야로 파악되고 있지 않는 등 인터넷 경제의 규모는 기존 통계로는 파악하기 어려움

○ 따라서 인터넷 경제 측정 방법론 연구를 통한 인터넷 관련 기업(비즈니스)의 경제규모에 대한 측정은 빅데이터 환경 변화와 4차 산업혁명 시대에 부합하는 통계 생산이라는 관점에서도 의의가 있음

4. 연구 범위와 연구 방법

가. 연구 범위

○ (연구범위) 인터넷 경제 측정기법 연구는 연구 대상을 인터넷 경제 및 플랫폼 경제로 정하고,

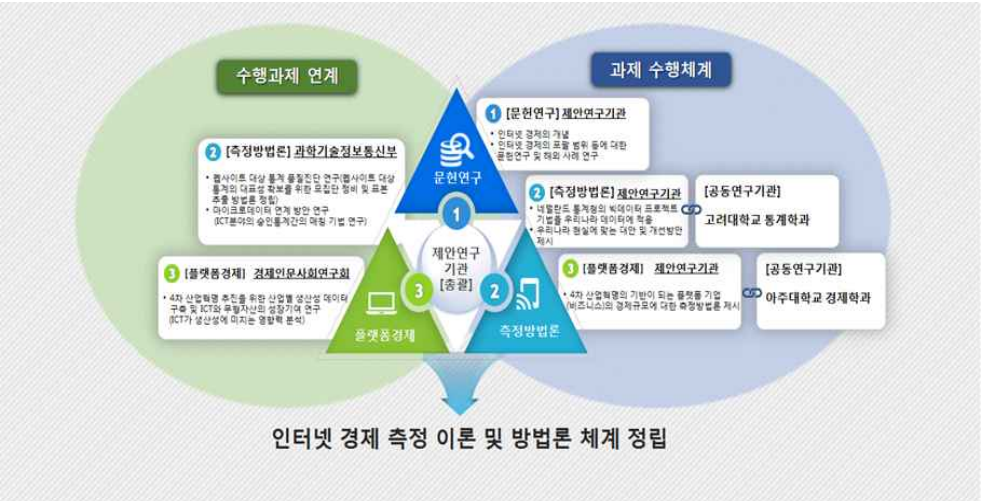
- 기존문헌 및 해외사례 연구, 네덜란드(CBS)의 인터넷 경제 측정 방법론 및 플랫폼 경제 측정방법론을 적용, 우리나라 통계청의 공공데이터와 민간데이터를 활용하여 시산하는 것으로 특정하며
- 네덜란드 방법론에 대한 한계점 및 대안을 검토하여 우리나라 상황에 맞는 인터넷 경제 측정방법을 제시하고자 함

○ (특징·장점) 인터넷 경제 측정 전문가와 공동연구, 제안기관이 수행 중인 관련 연구의 내용과 연계

- (공동연구 수행) 인터넷 경제 측정 이론 및 방법론 분야의 전문성을 보유한 대학과 공동연구 수행을 통해 연구의 품질 향상 도모
- (수행 과제 연계) 제안 기관이 2018년 진행 중인 웹사이트 대상 통계 품질진단 연구(웹사이트 대상 통계의 대표성 확보를 위한 모집단 정비 및 표본 추

출 방법론 정립), 마이크로데이터 연계 방안 연구(ICT분야의 승인통계간의 매칭 기법 연구), 4차 산업혁명 추진을 위한 산업별 생산성 데이터 구축 및 ICT와 무형자산의 성장기여 연구(ICT가 생산성에 미치는 영향력 분석)와 연계하여, 본 연구의 품질제고 도모

〔그림 1〕 과제 수행체계



나. 연구 방법

○ 인터넷 경제 측정 방법론 개선 방안 제시

– 인터넷 경제 측정 이론 및 방법론 연구를 관련 문헌 연구와 해외사례 연구, 네덜란드 통계청의 인터넷 경제 측정 방법론 적용 연구, 플랫폼 경제 측정 방법론 연구로 구분하여 분석하고, 이를 기초로 인터넷 경제 측정 이론 및 방법론 체계 정립의 기본 방향을 제시함

– 인터넷 경제 측정과 관련된 학계, 기업(McKinsey), 국제기구(OECD) 등의 관련 연구와 해외사례를 벤치마킹하여 측정 방법론 개선 방안을 제시

〈표 1〉 인터넷 경제 측정 방법론 정립 기본 방향

인터넷 경제 측정 이론 및 방법론 체계 정립

세부 연구 과제	주요 내용
1. 인터넷 경제, 플랫폼 경제 문헌 연구	인터넷 경제의 개념, 포괄 범위 등에 대한 문헌연구 및 해외 사례 연구
2. 네덜란드 통계청의 인터넷 경제 측정 방법론 적용 연구	네덜란드 통계청의 빅데이터 프로젝트 기법 중 하나인 인터넷 경제 측정방법을 우리나라 데이터에 적용하고 우리나라 현실에 맞는 대안 및 개선방안 제시
3. 플랫폼 경제 측정방법론 연구	4차 산업혁명의 기반이 되는 플랫폼 기업(비즈니스)의 경제규모에 대한 측정방법론 제시

○ 우리나라 상황에 적합한 플랫폼 경제 측정방법론 제시

– 인터넷 경제/플랫폼 경제 측정 방법론을 구체화하기 위한 전략 체계 제시

〈표 2〉 인터넷 경제 측정 연구 수행 전략

기본방향	1 인터넷 경제 측정방법론 개선 방안 도출 2 플랫폼 경제 측정방법론 제시
세부방향	1-1. 인터넷 경제 측정 문헌 및 해외 사례 연구 1-2. 네덜란드 통계청 측정 방법론 적용 연구 1-3. 우리나라 현실에 부합하는 측정 개선방안 연구 2-1. 플랫폼 경제 측정을 위한 해외사례 연구 2-2. 플랫폼 경제 측정방법론 연구

5. 주요 연구 결과

가. 인터넷(플랫폼) 경제 관련 기존 문헌 및 해외사례 연구

○ 인터넷 경제의 개념

– OECD(2008)에서 인터넷 경제란 인터넷과 정보통신기술에 기반한 모든 경제·사회·문화적 활동으로 정의

– OECD(2015)는 인터넷 경제 대신 디지털 경제라는 용어로 변경, 디지털 경

제를 인터넷상에서 전자적 상거래를 통해 재화나 서비스의 거래를 촉진하는 디지털 기술에 기반 한 시장으로 정의

- BCG(2010)는 인터넷 경제의 파급 효과 분석을 위한 범위를 GDP내에서 파악 가능한 범위로 한정
- McKinsey(2011)의 경우 명확히 인터넷 경제를 정의하기 보다는 인터넷과 관련된 활동들을 인터넷 서비스와 네트워크의 창조 및 사용과 관련된 활동이라고 포괄적으로 규정

〈표 3〉 인터넷 경제, 디지털 경제, 플랫폼 경제의 개념 비교

인터넷 경제	○ 인터넷과 관련된 정보통신기술에 기반 경제, 사회, 문화활동의 전 영역으로 정의
디지털 경제	○ 인터넷상에서 전자적 상거래를 통해 재화나 서비스가 거래되는 촉진하는 디지털 기술에 기반한 시장으로 정의
플랫폼 경제	○ 양면 시장, 온라인 상거래, 공유 경제 등 플랫폼에 의해 실현된 모든 경제・사회적 활동으로 정의

○ 인터넷 또는 디지털 경제 시장 규모

- 주요 민간 컨설팅업체는 인터넷 또는 그 유사 활동들이 GDP에 기여하는 정도를 산출
- McKinsey(2011)는 2009년 기준 전세계 인터넷 기여도 규모가 1조 6,720억 달러이며, 전세계 GDP의 2.9%를 차지할 것으로 추정
- BCG(2012)는 2010년 기준 G20 국가의 인터넷 경제 규모가 약 55조 달러에 이르며, 전세계 GDP의 4.1%를 차지할 것으로 추정
- OECD(2017)는 디지털 경제 시장 규모를 전체 산업내 ICT 부가가치 비중으로 산출해 OECD 평균 전체 산업내 ICT 부가가치는 2015년 기준 5.4%임

○ 인터넷 경제 측정 방법론에 대한 사례 연구

- 인터넷 경제 규모 측정을 위한 주요 방법론은 크게 산출 방식에 따른 방법론과 산출 범위에 따른 방법론으로 구분

– 산출 방식에 따른 방법론은 총 지출 기반의 접근 방식과 총 소득 기반의 접근 방식으로 구분

– 산출 범위에 따른 구분은 직접적 효과, 동태적 효과, 간접적 효과로 구분

○ 인터넷 경제 규모 측정 사례 연구 결과 전통적 방식에 한계가 존재하기 때문에 인터넷 경제를 구성하는 단위를 경제주체별 인터넷 활용 방식 등으로 새롭게 분류하고 기존 통계에서 노출되지 않는 인터넷의 기여분을 측정할 수 있는 새로운 방식 필요

나. 인터넷 경제 측정 방법론 연구

(1) 네덜란드 통계청의 인터넷 경제 측정 방법론 적용 연구

○ (연구 목적) 인터넷 경제의 중요성에 대한 이해도 증진

– (세부 목표) 인터넷 경제의 정의를 실체 확인과 빅데이터 활용 가능성이라는 실용적인 관점에서 접근, 네덜란드 인터넷 경제의 규모와 전체 경제에서 차지하는 비중과 중요성을 탐색, 새로운 측정 방법론의 가능성을 모색, 현재 통계의 개념, 표준 분류와의 차이점을 탐구

○ (분류 체계) 웹데이터 크롤링 전문회사인 Dataprovider의 데이터를 활용, 네덜란드 정부와 산업계, 학계, 구글이 협력, 분류체계 도출

※ 기업체가 인터넷을 어떻게 사용하는지에 따라 웹사이트 미보유(A), 웹사이트 보유(B), 온라인 상점(C), 온라인 서비스(D), 인터넷 관련 ICT(E) 기업체로 분류

※ 분류 B~E를 인터넷 경제 활동 기업으로 정의하고, 이중 C~E를 인터넷 경제의 핵심(core)으로 분류

○ (활용 데이터) 민간회사의 웹사이트 크롤링 데이터와 네덜란드 통계청의 경제 데이터를 활용

※ 웹사이트 크롤링 데이터는 웹사이트명, 업체명, 사업자등록번호(CoC number), 접속 횟수 등으로 구성되어 있으며 이 외에 주문, 구매, 쇼핑 카트, 예약 등의 이용자 정보 데이터베이스로 구성

※ 네덜란드 통계청의 경제 데이터로는 기업 규모(종업원 수), 분야, 업력 등으로 구성된 기업등록부(General Business Register, GBR)와 재무 정보로 구성된 생산통계(Production Statistics)를 활용

※ 이 외에 부가가치세 데이터, 매출, 지역 통계, 행정자료, 소매조사(Retail Survey), ICT 활용 조사 등을 활용

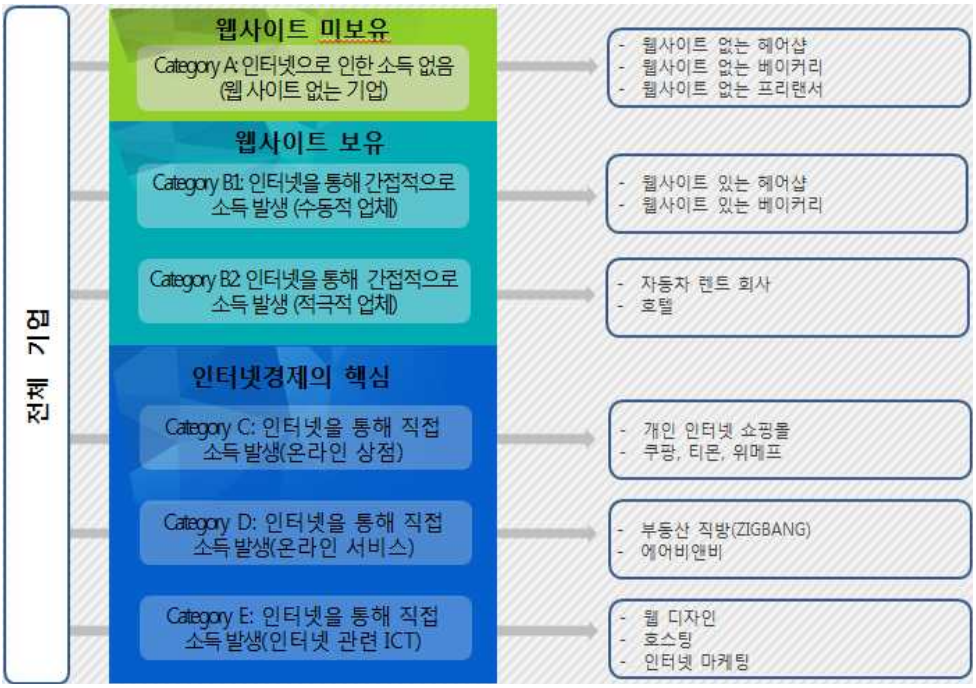
○ (데이터 연계 방법론) 민간회사의 웹사이트 데이터를 네덜란드 통계청의 사업자등록번호(CoC), 사업체, 기업체 데이터와 연계

○ (웹 사이트 연계방식) 웹사이트를 사업자등록번호와 연계하고 이를 사업체, 기업체 단위로 재병합하여, Dataprovider에서 제공한 250만개의 웹사이트가 55만개의 기업체 데이터로 병합

○ (측정 결과) 2015년 기준 36%의 기업이 웹사이트 보유, 웹사이트를 보유하지 않은 기업의 83%가 자영업자

– 코어인터넷 기업(온라인 상점, 온라인 서비스, 인터넷 관련 ICT기업)은 5만개로 종사자는 345,000명(4.4%), 매출은 1040억 유로(7.7%)로 주로 건설, 숙박 및 음식, 교통 분야

〔그림 2〕 인터넷 활용에 따른 기업 분류



출처: CBS(2016), Measuring the internet economy in The Netherlands: a big data analysis 재구성

(2) 우리나라 데이터 적용 분석

○ (데이터 입수) 데이터프로바이더로부터 우리나라 웹사이트의 크롤링 데이터를 구입하고 통계청의 기업등록부 자료를 입수

– 데이터프로바이더의 한국 웹사이트 크롤링 데이터는 925,961개

– 데이터프로바이더 데이터는 6개 섹션 128개 변수로 구성

○ 통계청의 경제 데이터와 연계를 위한 key 변수를 파악하고 데이터프로바이더 정보를 이용하여 개별 웹사이트를 “인터넷 사용에 따른 업체분류”인 카테고리 A~E를 부여

○ (데이터 연계 방법론) 네덜란드 방법론을 차용하여 웹사이트 데이터를 한국 통계청 기업등록부와 연계

○ (매칭) 사업자등록번호를 이용해 데이터프로바이더의 데이터와 기업등록부를 m:m 매칭했고, 기업등록부의 11,434개 레코드가 매칭

○ (분석단위) 데이터 매칭 및 주요통계 추출 과정에서는 기업등록부 데이터의 특성상 통계청의 고유번호(기업등록부 id)를 분석단위로 이용

○ 데이터 프로바이더의 한국 데이터 특성

– 사업체명이 있는 웹사이트는 14만 6천 개로 전체의 16%

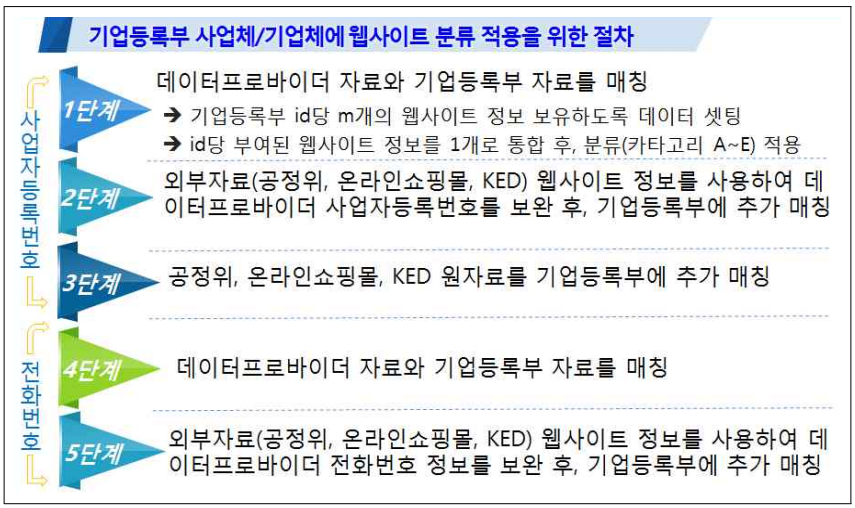
– 사업자 등록번호가 기입된 웹사이트는 1만 3천 개로 전체의 1%에 불과

– 전화번호가 기입된 웹사이트는 27만 개로 전체의 30% 수준

○ 우리나라 데이터 적용을 위한 수정된 매칭 전략

– 데이터 프로바이더의 정보를 기업등록부에 매칭하기 위해 다양한 외부 정보(공정위, 온라인쇼핑몰, KED)를 활용한 수정된 매칭 전략을 활용함

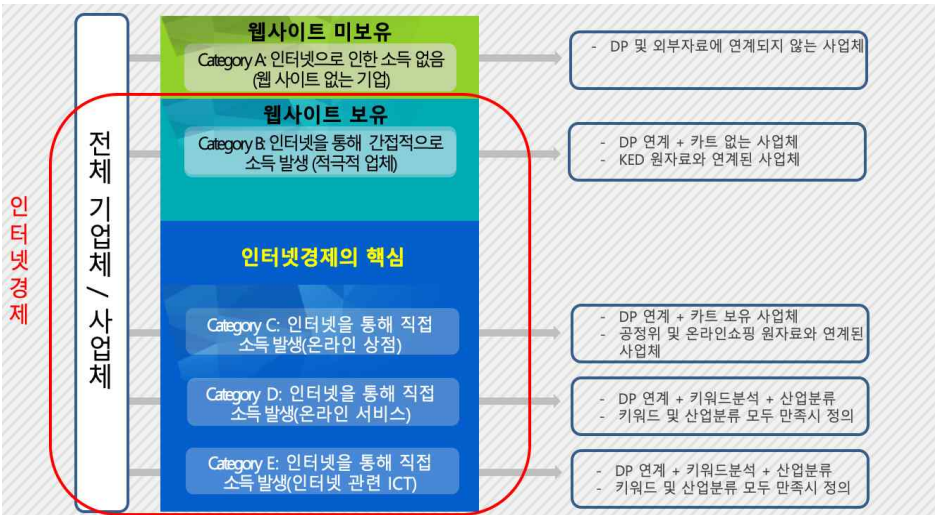
〔그림 3〕 우리나라 데이터 적용 분석을 위한 매칭 전략



○ 카테고리 D와 E의 키워드 분석

– 네덜란드 인터넷 경제 분류의 D와 E에 해당하는 국내 업체를 판별하기 위해 해당 카테고리와 연관된 키워드 분석을 위해 국내 주요기업과 네덜란드 데이터간의 매칭 작업을 수행

〔그림 4〕 인터넷 사용에 따른 업체 분류



(3) 우리나라 인터넷 경제 측정 결과

○ 네덜란드(CBS)의 인터넷 경제 측정 방법론을 우리나라 통계청 데이터와 다양한 공공, 민간데이터에 적용하여 시산한 결과는 다음과 같음

○ 데이터 단계별 매칭 결과

– 데이터프로바이더가 크롤링한 데이터와 이를 보완하기 위해 외부 자료(공정위 “통신판매사업자 신고자료”, 통계청 “온라인쇼핑동향조사 모집단 자료”, “KED 웹사이트자료)를 활용하여 통계청의 기업등록부(BR)와 매칭

– 추가 데이터를 활용함에 따라 사업자등록번호만을 사용한 단계에 비해 최종단계의 매출은 약 2.5배 증가

〈표 4〉 데이터 매칭 단계별 매출액 및 종사자 수

구분	매출액 (조원)	비중 (%)	종사자수 (명)	비중 (%)
기업등록부(BR) 전체	5,892	–	20,430,684	–
1단계(사업자등록번호 기준, DP → BR)	1,214	20.6	1,873,121	9.2
2단계(사업자등록번호 기준, DP + 외부자료 → BR)	2,682	45.5	5,855,072	28.7
3단계(사업자등록번호 기준, 외부자료 → BR)	2,720	46.2	5,974,356	29.2
4단계(전화번호 기준, DP → BR)	2,828	48.0	6,423,199	31.4
5단계(전화번호 기준, DP + 외부자료 → BR)*	2,837	48.2	6,467,073	31.7

○ 연계 데이터 출처별 매칭 결과

– 기업등록부와 연계된 데이터 출처별 기업체 및 사업체 수 매칭 결과는 다음과 같음

〈표 5〉 연계 데이터 출처별 기업체 및 사업체 수

구분	기업체수 (개)	비중 (%)	사업체수 (개)	비중 (%)
기업등록부와 연계된 총 관측치	216,099	100.0	313,271	100.0
(1그룹) 사업자등록번호 기준 BR-DP	122,390	56.6	203,371	64.9
BR－DP 원자료	8,178	3.7	28,483	9.1
BR－외부자료 통해 보완한 DP자료	114,212	52.9	174,888	55.8
(2그룹) 사업자등록번호 기준 BR－외부자료	32,214	14.9	37,308	11.9
(3그룹) 전화번호 기준 BR-DP	61,495	28.5	72,592	23.2
BR－DP 원자료	50,318	23.3	58,133	18.5
BR－외부자료 통해 보완한 DP자료	11,177	5.2	14,819	4.7

－ 연계된 데이터 출처별로 기업체 기준 매출액 및 종사자 수는 다음과 같음

〈표 6〉 연계 데이터 출처별 매출액 및 종사자 수

구분	매출액 (조원)	비중 (%)	종사자 수 (명)	비중 (%)
기업등록부와 연계된 총 관측치	2,837	100.0	6,467,073	100.0
(1그룹) 사업자등록번호 기준 BR-DP	2,682	94.5	5,855,072	90.5
BR - DP 원자료	1,214	42.8	1,873,121	29.0
BR - 외부자료 통해 보완한 DP자료	1,468	51.7	3,981,951	61.5
(2그룹) 사업자등록번호 기준 BR－외부자료	37	1.3	119,284	1.8
(3그룹) 전화번호 기준 BR-DP	118	4.2	492,717	7.7
BR - DP 원자료	109	3.8	448,843	7.0
BR - 외부자료 통해 보완한 DP자료	9	0.4	43,874	0.7

○ 인터넷 경제 속성별 매칭 결과

－ (분포) 인터넷 경제 분석에 활용한 최종 데이터는 기업체 기준으로 총 216,099개이며, 그룹별 분포를 살펴보면 B그룹 46%(99,391개), C그룹 53%(114,619개), D그룹 0.9%(1,916개), E그룹 0.1%(173개)로 나타났음

－ (매출액) 인터넷 경제에 속한 기업체의 총 매출액은 2,837조 원이며, B그룹

이 70.9%(2,011조 원), C그룹이 28.6%(811조 원), D그룹이 0.5%(13조 원), E그룹이 0.1%(2조 원)을 차지하는 것으로 나타났음

- (종사자) 인터넷 경제에 속한 기업체의 총 종사자 수는 6,467,073명이며, B그룹이 70.8%(4,580,103명), C그룹이 28.3%(1,832,580명), D그룹이 0.7%(46,359명), E그룹이 0.1%(8,031명)을 차지하는 것으로 나타났음
- (지역별 분포) 기업등록부내 인터넷 경제 기업의 경우 서울(36.8%), 경기도(26.3%), 부산(5.0%) 순으로 분포가 높았는데 인터넷 기업이 서울과 경기 등 수도권에 밀집된 정도가 상대적으로 높게 나타남
- (산업별 분포) 기업등록부내 인터넷 경제 기업의 경우 도매 및 소매업(34.3%), 제조업(21.4%), 출판, 영상, 방송통신 및 정보서비스업(8.1%) 순으로 분포가 높은데 전체 기업대비 도매 및 소매업의 비중이 높으며, 특히 제조업의 비중이 높은 것이 특징

〈표 7〉 인터넷 경제 속성별 기업체 수, 매출액 및 종사자 수

구분	기업체수	분포(%)	매출액(조 원)	분포(%)	종사자수(명)	분포(%)
인터넷 경제 기업	216,099	100.0	2,837	100.0	6,467,073	100.0
B그룹	99,391	46.0	2,011	70.9	4,580,103	70.8
C그룹	114,619	53.0	811	28.6	1,832,580	28.3
D그룹	1,916	0.9	13	0.5	46,359	0.7
E그룹	173	0.1	2	0.1	8,031	0.1

○ 인터넷 경제 속성별 주요 변수 분포

- 기업등록부 전체 기업을 대상으로 인터넷 단계/종사자 규모, 인터넷 단계/업력, 인터넷 단계/지역, 인터넷 단계/산업으로 구분하여 분포를 분석

※ 인터넷 경제 기업을 나타내는 B~E그룹에 포함되지 않은 모든 기업은 A로 정의

- (종사자규모) A그룹의 경우 0~1명 종사자를 보유한 소규모 기업체의 비중이 85.8%를 차지하며, C그룹(64.5%)과 D그룹(67.3%)도 소규모 기업체 비중이 높았으나 B그룹과 E그룹은 상대적으로 고른 분포를 보임

– (업력) B그룹과 E그룹의 경우, 10년 넘게 영업한 기업체의 수가 많았고, A그룹, C그룹, D그룹은 업력이 5년 이하인 기업체가 해당 그룹에 속한 기업의 절반 이상을 차지

– (지역) A그룹은 대부분의 지역에 분포, B그룹과 C그룹은 서울과 경기도, E그룹은 서울 지역에 대한 비중이 높았으며, D그룹의 경우 서울, 경기도, 강원도, 경상남도, 충청남도 등에 주로 위치한 것으로 나타남

– (산업) B, C그룹의 경우 제조업, 도매 및 소매업, 출판,영상,방송통신 및 정보서비스업이, D그룹은 숙박 및 음식점업, 사업시설 관리 및 사업지원 서비스업이, E그룹의 대부분은 출판,영상,방송통신 및 정보서비스업으로 나타남

(4) 네덜란드 방법론에 대한 한계점

○ 연구의 한계

– 네덜란드 측정방법론에 따른 인터넷 경제에 대한 규모 추정은 경제 내의 생산구조를 파악하는데 기여할 것으로 보이나,

– 네덜란드 측정 방법론은 한 국가의 인터넷 경제 규모를 측정하기 위한 파일럿 연구의 성격이 큼

– 따라서 우리나라 현실에 적합한 인터넷 경제 규모 측정을 위해서는 글로벌 경쟁력을 통해 대규모의 수출이 발생하고 있는 ICT 기기 부문에 대한 고려가 필요

○ 인터넷 경제/플랫폼 경제 규모에 대한 측정방법론 검토

– 인터넷 경제의 측정을 위해서는 이를 측정할 수 있는 조작적 정의 및 범위 설정을 위한 정량적 측정이 요구됨

– (대안1) 크롤링 데이터 활용: 네덜란드 통계청에서 적용한 방법론을 이용하되, 국내의 웹크롤링 전문 업체를 통해 크롤링 데이터를 확보하여 통계청 데이터를 결합, 인터넷경제 규모를 추정

– (대안2) 기존 통계자료 활용: 인터넷경제와 플랫폼 경제의 정의 및 포함관계 분석을 통해, 광고시장 규모를 통한 플랫폼경제 규모를 간접 추정하거나 기존 정보화통계조사 등을 통해 산업별 “네트워크에 연결된 컴퓨터 보

유 사업체” 비율 등을 활용해 인터넷경제의 규모를 추정

– 인터넷경제와 플랫폼경제의 다면적인 현황을 파악하기 위해서는 수요, 공급, 인프라, 이용자 등 다양한 관점의 통계조사 결과 활용이 필요

다. 플랫폼 기업(비즈니스) 경제 규모에 대한 측정방법론 연구

○ 디지털 경제, 인터넷 경제, 플랫폼 경제의 개념과 특징

– 플랫폼 경제는 운영체제, 검색, SNS, 거래중개 등의 핵심적 플랫폼의 역할에 기초하여 판매자, 이용자, 서비스 제공자들이 상호작용을 통해 조직화된 디지털 또는 인터넷 생태계를 의미

– 플랫폼 경제란 광의의 의미로는 인터넷경제(디지털경제)를 의미, 협의의 의미로는 인터넷 플랫폼을 통한 서비스 제공(카테고리 D)

○ 양면시장(two-sided market)이란 두 그룹의 이용자(또는 사업자)가 특정한 플랫폼을 통해 상호작용하여 새로운 가치를 창출하는 시장

○ 플랫폼 경쟁의 유형에 따른 시장구조 특성

– 플랫폼의 양측 이용자가 모두 싱글호밍인 경우 (결혼정보회사)

– 플랫폼의 한 측 이용자는 싱글호밍, 다른 측 이용자는 멀티호밍인 경우 (방송, 인터넷포털)

– 플랫폼의 양측 이용자가 모두 멀티호밍인 경우 (호텔예약 사이트, 인터넷 쇼핑몰)

라. 시사점

○ 본 연구는 인터넷경제 규모 측정을 위해 빅데이터 방법(크롤링)을 통해 수집한 웹사이트 데이터와 통계청의 기존 산업통계를 결합하는 방법을 활용한 파일럿 성격의 연구임

– 크롤링 데이터의 품질, 매칭을 위한 연계키, 인터넷 경제 카테고리 분류를 위한 보조 정보의 부재, 매칭 방법론의 정합성 등의 측면에서 판단할 때, 전체 경제에서 차지하는 인터넷 경제의 규모를 정확하게 산출하기에는 시기상조인 것으로 판단함

– 신뢰할 수 있는 인터넷 경제 규모 측정 결과를 산출하기 위해서는 웹사이트

모집단 정보의 확보, 웹크롤링 결과의 품질 제고, 해당 기업체 단위의 온라인 및 오프라인 매출 비중을 추정하기 위한 보조 정보의 확보 등이 필수적이며 다음과 같은 추정을 위한 여건과 관련 기초 연구가 향후 필요할 것으로 판단됨

○ (웹사이트 모집단 정보) 인터넷 경제 측정의 출발은 웹사이트 모집단 정보 확보임

- 대안 1: 통계청의 주요 조사(전국사업체조사, 광업제조업조사 등) 시 사업체(기업체)별 웹사이트 조사
- 대안 2: 한국인터넷진흥원 도메인 등록시 필요 정보(업체명, 사업자등록번호 등) 기입 의무화 방안. 단 기존 등록 정보는 결측 항목 조사 등 정비가 필요(단점은 .kr 이외 .com 등 도메인 정보는 확보 곤란)

○ (웹크롤링) 통계청 BR데이터와 매칭 성공률을 높이기 위해서는 필요 정보(사업체명, 사업자 등록번호, 전화번호, 쇼핑카트 유무, 취급 품목 등) 수집을 위한 크롤링의 성능이 매우 중요. 기술력이 확보된 국내 업체와 협업 또는 자체적으로 크롤링 능력 확보가 필수

- 기존 통계나 행정자료가 포괄하지 못하는 기업 특성을 보완하는 것이 크롤링 등 빅데이터 방법을 사용하는 주된 이유임을 감안할 때, 기존 자료에서 미흡한 영역의 정보를 확보하기 위한 방법론 개발이 중요
- 한국어로 구성된 홈페이지 분석 능력이 중요하므로, 웹크롤링에 대한 충분한 기술력을 보유한 국내 업체를 통해 연구 수행 가능성을 모색할 필요가 있음

○ (인터넷경제 카테고리 분류) 크롤링 데이터를 기초로 인터넷경제 카테고리를 자동 분류하는 판별 모형의 성능이 매우 중요하므로 이를 위한 빅데이터 방법론 개발이 필수

〈표 8〉 인터넷경제 카테고리 분류별 판정 난이도

구분	판단 기준	난이도
B	웹사이트 유무	낮음
C	쇼핑 카트 유무	보통
D	특정 취급 품목(서비스) 유무	매우 높음
E	특정 취급 품목(서비스) 유무	높음

○ (기초 연구) 카테고리 분류(특히 D, E)시 활용한 키워드 판정 방법론의 성능 검증, 인터넷 경제 규모 추정 결과 신뢰도 측정 등 관련 기초 연구 필요

○ (기업 정보) 인터넷경제의 특성을 고려하여 정확한 경제 규모를 산출하기 위해서는, 기업의 매출액 등 주요 지표를 온라인 기반과 오프라인 기반으로 나눌 수 있는 보조 자료가 필요

– 기업등록부 등 행정자료에서는 기업의 총 매출액에 대한 정보만 제공되고 있음을 고려할 때, 이를 보완할 수 있는 자료 구축이 필요

※ 웹사이트를 보유하였으나 온라인 매출 규모가 미미한 기업의 총 매출액을 인터넷경제 규모 추정에 포함할 경우 과대추정 가능

○ (한국형 측정 방법) 본 보고서의 분석은 네덜란드 통계청의 방법론을 기반으로 한 것이므로, 한국의 산업 특성을 반영한 인터넷경제 측정 방법을 발전시켜 나가야 함

– 한국의 인터넷경제의 특성을 다른나라와의 국제비교를 통해, 한국 특성에 부합하는 고유 방법론을 구축하는 것이 필요

– 오픈마켓 등 인터넷경제에서 중요한 대상을 포괄할 수 있는 측정 방법 필요

– 인터넷경제를 파악함에 있어 전통적인 산업 분석에서는 고려하지 않았던 지표 및 특성이 중요하므로, 다양한 관점의 통계지표 개발을 위한 접근이 요구됨

※ 예를 들어 인터넷 경제 측정에 있어서, 전통적인 사업체의 지리적 소재지 분포는 큰 의미를 갖지 못할 가능성이 있음

6. 기대효과 및 활용방안

○ (학술적 효과) 거시 및 산업 데이터인 국민계정(SNA) 데이터와 산업분류를 이용한 기존의 학술 연구와 달리 네덜란드 측정방법론은 미시데이터를 연계하는 방식과 빅데이터 방법론을 활용한 방법론의 제시를 통해 학술적 측면의 새로운 방법론 적용에 따른 연구 진작과 다양한 시사점을 도출할 것으로 보임

○ (정책 효과) 파일럿 프로젝트 과정에서 도출된 다양한 문제점을 기초로 정확한 인터넷 경제 규모 측정을 위해 필요한 모집단 정보 확보, 온/오프라인 매출 분리를 위한 보조 정보의 확보 등에 대한 필요성 제시를 통해 향후 관련 정책 지원을 위한 기초 데이터 정비 방안 수립에 기여

제 1 장 서 론

제 1 절 연구의 목적 및 필요성

1. 연구의 목적

○ 본 연구의 목적은

– 첫째, 인터넷 경제(디지털 경제) 측정에 대한 기존 연구와 해외사례를 분석하고

– 둘째, 네덜란드 통계청(CBS)의 인터넷 경제 측정 방법론을 우리나라 통계청 데이터를 적용하여 시산하며

– 셋째, 네덜란드 방법론에 대한 한계점 및 대안을 검토하여 우리나라 상황에 맞는 인터넷경제 측정방법을 제시

2. 연구의 필요성

○ 최근 인터넷, ICT 등의 발달로 경제 환경이 산업경제에서 인터넷 경제 또는 디지털 경제로 빠르게 전환하는 가운데 디지털 경제의 비중이 갈수록 높아지는 추세

– 현재 인터넷의 역할은 단순한 커뮤니케이션 수단을 넘어 다양한 경제활동에 활용되는 핵심 인프라로 자리매김 하였으며 그 가치는 빠르게 증대되고 있음

– 시간과 장소의 제약 없이 다양한 기기에서 인터넷에 접속이 가능해짐에 따라 인터넷을 통해 제공되는 서비스가 매우 다양화되고 기존 상거래 방식도 인터넷을 통해 빠르게 대체되고 있음

○ 인터넷은 기본적으로 컴퓨터 네크워크로서 모든 컴퓨터를 하나의 통신망으로 연결해주는 글로벌 네크워크를 의미하며, 인터넷 이용자수, 인터넷 속도 등 인터넷 기술의 단순한 활용을 측정하는 것과 인터넷 경제를 측정하는 것은 차이가 있음

- 일반적으로 경제 활동이란 인간의 생활에 필요한 재화나 용역을 생산·분배·소비하는 모든 활동을 의미
- 인터넷 경제는 인터넷이라는 인프라에 기반하여 혹은 그것을 활용한 재화나 서비스의 생산, 유통, 판매, 소비 등 다양한 경제 활동들을 포괄한다고 할 수 있음

○ 인터넷의 확산으로 경제전반에 미치는 영향력이 확대됨에 따라 향후 적절한 정책수립을 위해 인터넷 경제의 규모와 경제적 파급효과를 추정하기 위한 방법론 구축의 필요성이 확대되고 있음

○ 인터넷 경제를 계량적으로 측정할 경우 인터넷 경제를 인터넷이라는 통신수단과 관련된 모든 산업 활동과 인터넷에 기반하여 수행되는 모든 경제활동을 포함할 것인지 아니면 인터넷을 통해 거래되는 재화나 서비스 또는 인터넷을 통한 금융적 거래활동만을 포함할 것인지에 대한 개념 정립이 매우 중요

- OECD는 인터넷 경제의 일관성과 비교 가능성을 확보하기 위하여 인터넷 경제의 측정은 계량화 할 수 있는 인터넷의 다양한 경제적 효과를 포함하여야 한다는 의견을 제시(OECD, 2011)
- 인터넷 경제의 측정을 위한 접근 방법을 1) 인터넷 경제의 직접적 효과, 2) 동태적 효과, 3) 간접적 효과로 구분하고 이를 측정하기 위한 시도를 진행(OECD, 2012)

○ 인터넷 경제가 이미 산업 전 분야에 확산된 상황을 감안할 때 국내 인터넷 경제의 규모 및 파급효과를 파악하기 위해 종합적인 연구가 필요

- 또한 4차 산업혁명의 기반이 되는 플랫폼 경제(platform economy)에 대한 중요성이 부각되고 있어 정책 지원을 위한 플랫폼 경제의 현황 파악과 분석의 중요성이 부각
- 인터넷 경제의 중요성에도 불구하고 그 규모를 측정한 사례가 드물고, GDP에서도 독립된 분야로 파악되고 있지 않는 등 인터넷 경제의 규모는 기존 통계로는 파악하기 어려움

○ 인터넷 경제의 확산은 데이터 생산 관점에서 볼 때 정확한 거래 데이터가 적시에 양산된다는 점에서 기존 통계 조사를 보완, 대체 가능하다는 점에서 주목받

고 있음

- 특히 국가승인통계에서 표준방법으로 인정하고 있는 면접조사의 경우 원천적 접촉 불가능(보안시스템 강화 공동주택), 조사 거부(프라이버시 보호 등), 접촉 곤란(1인 가구, 특정 직업/연령대 등) 등으로 많은 어려움을 겪고 있는 실정임(정용찬 외, 2016)
- 비용 부담이 큰 인구주택총조사의 대안으로 행정자료를 활용하거나 기존의 행정자료와 조사 자료를 결합하는 방식으로 데이터의 효용을 증대시키기 위한 다양한 시도가 추진 중

○ 해외 주요국의 경우 물가통계, 에너지통계, 교통통신통계 등 다양한 분야에서 국가통계에 빅데이터를 활용하기 위한 시도가 진행 중

〈표 1－1〉 빅데이터 국가통계 활용 해외사례

구분	데이터 유형	주요 내용
교통통계	센서 데이터 (Sensor data)	－ 네덜란드는 도로 상에 설치한 2만여 개의 차량감지기를 통해 차종별 교통량, 속도 등의 정보 수집 － 경기 변동과 관련 있는 화물차량의 물동량 정보 분석, 지역별 통행량지수(traffic index) 산출
소셜미디어 통계	SNS, 블로그	－ 네덜란드 통계청은 소셜미디어 메시지에 대한 감성(sentiment) 분석 － 감성분석 결과 경기와 상관관계가 높은 것으로 나타나 주간소비자신뢰지수 산출에 활용 가능 － UN Global Pulse의 트위터와 물가 관계 분석
물가통계	온라인 데이터	－ Eurostat는 인터넷의 물가 정보를 자동 수집, 소비자물가지수(CPI) 산출에 참고 － MIT의 Billion Prices Project － 중국 통계청은 2013년 11월 알리바바, 바이두 등 11개 전자상거래업체와 빅데이터 물가지수 개발 위한 파트너쉽 협약 체결

자료: UNECE(2013), What does “Big Data” mean for official statistics? Peter Struijs(2014), Traffic loop data for transport statistics, http://bpp.mit.edu/, 정용찬(2016) 재인용

○ 국가통계에서 빅데이터의 활용은 기업통계, 개별 산업별 통계에서 인구통계, 실업통계 등 공공부문 통계로 확대되는 추세(정용찬 외, 2017)

〈표 1－2〉 빅데이터 국가통계 활용 분야 확대 사례

구분	데이터 유형	주요 내용
물가통계	온라인 데이터 (U.S.CensusBureau)	－ 표본조사 데이터와 온라인 거래 데이터의 연계 － 자동차부품, 보석/시계 POS데이터 분석 － 소지역추정량 산출 모형 개발
인구통계	모바일폰 데이터 (Eurostat, 벨기에통계청)	－ 모바일폰 데이터와 상주인구통계 비교분석 － 인구, 관광, 교통, 이사 통계에 활용 가능
실업통계	온라인 데이터 독일통계청	－ 구인구직사이트((job portal)의 데이터를 활용, 일자리 추정
관광통계	교통량/모바일폰/ 신용카드데이터 (스페인통계청)	－ 빅데이터와 조사 데이터간의 비교 분석 (국적, 관광 일수, 지출 규모 등)

자료: Eurostat, Q2016 European Conference on Quality in Official Statistics, 31st May – 3rd June, 정용찬 외(2017) 재인용

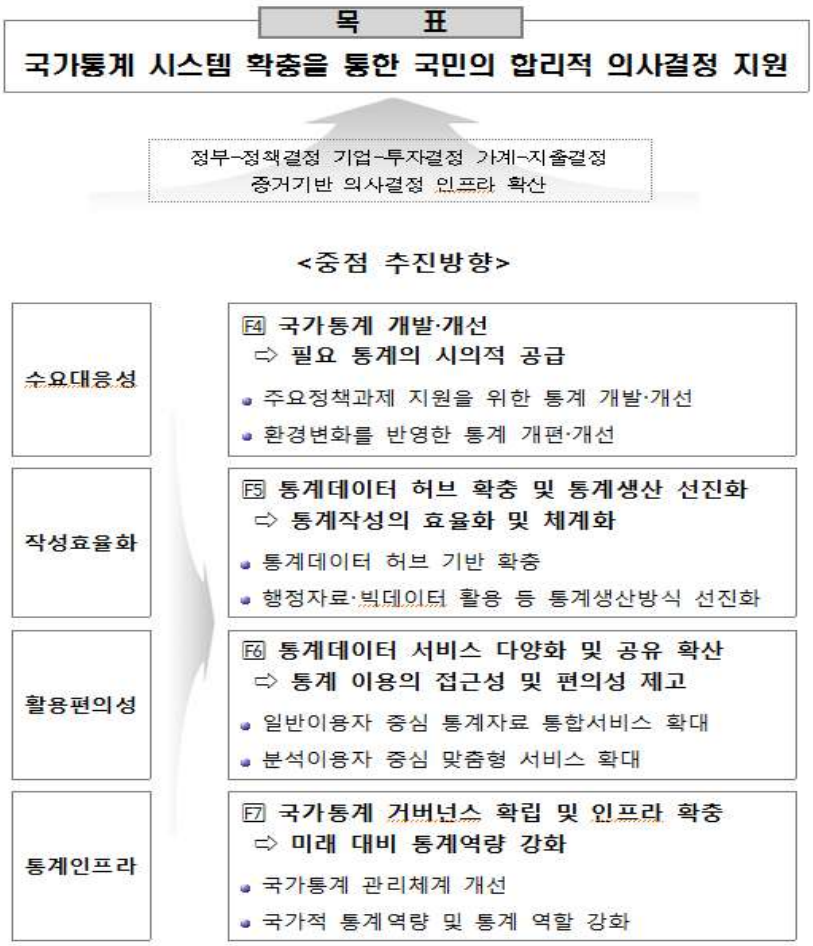
○ 최근 통계청은 ‘국가통계 시스템 확충을 통한 국민의 합리적 의사결정 지원’을 목표로 한 ‘제2차 국가통계발전 기본 계획(2018～2022)’을 발표

－ 통계 수요 대응성, 작성 효율화, 활용 편의성, 통계 인프라 확충을 위해 국가통계 개발·개선, 통계데이터 허브 확충 및 통계생산 선진화, 통계데이터 서비스 다양화 및 공유 확산, 국가통계 거버넌스 확립 및 인프라 확충을 4대 중점 추진과제로 제시

－ 특히 통계 데이터 허브 확충 및 통계 생산 선진화를 위해서 행정자료와 빅데이터 활용을 강조

○ 따라서 인터넷 경제 규모 측정 연구는 빅데이터 환경 변화와 4차 산업혁명 시대에 부합하는 통계 생산이라는 관점에서 매우 중요

〔그림 1-1〕 제2차 국가통계발전 기본 계획 추진방향



제 2 절 연구 범위와 연구 방법

1. 연구 범위

○ (연구범위) 인터넷 경제 측정기법 연구는 연구 대상을 인터넷 경제 및 플랫폼 경제로 정하고,

– 기존문헌 및 해외사례 연구, 네덜란드(CBS)의 인터넷 경제 측정 방법론 및 플랫폼 경제 측정방법론을 우리나라 통계청의 공공데이터와 민간데이터에

적용하여 시산하는 것으로 특정하며

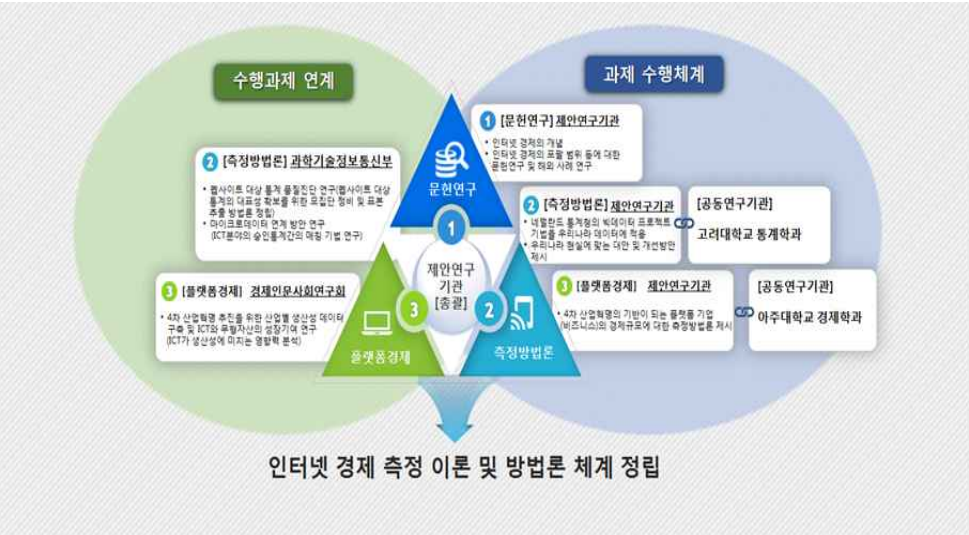
– 네덜란드 방법론에 대한 한계점 및 대안을 검토하여 우리나라 상황에 부합하는 측정방법을 제시하고자 함

○ (특징·장점) 인터넷 경제 측정 전문가 협의체 운영으로 연구 전반의 품질 향상 도모

– (공동연구 수행) 인터넷 경제 측정 이론 및 방법론 분야의 전문성을 보유한 대학과 공동연구 수행을 통해 연구의 품질 향상 도모

– (수행 과제 연계) 제안 기관이 2018년 진행 중인 웹사이트 대상 통계 품질진단 연구(웹사이트 대상 통계의 대표성 확보를 위한 모집단 정비 및 표본 추출 방법론 정립), 마이크로데이터 연계 방안 연구(ICT분야의 승인통계간의 매칭 기법 연구), 4차 산업혁명 추진을 위한 산업별 생산성 데이터 구축 및 ICT와 무형자산의 성장기여 연구(ICT가 생산성에 미치는 영향력 분석)와 연계하여, 본 연구의 품질제고 도모

〔그림 1–2〕 과제 수행체계



2. 연구 방법

○ 인터넷 경제 측정 방법론 개선 방안 제시

– 인터넷 경제 측정 이론 및 방법론 연구를 관련 문헌 연구와 해외사례 연구, 네덜란드 통계청의 인터넷 경제 측정 방법론 적용 연구, 플랫폼 경제 측정방법론 연구로 구분하여 분석하고, 이를 기초로 인터넷 경제 측정 이론 및 방법론 체계 정립의 기본 방향을 제시함

– 인터넷 경제 측정과 관련된 학계, 기업(McKinsey), 국제기구(OECD) 등의 관련 연구와 해외사례를 벤치마킹하여 측정 방법론 개선 방안을 제시

〈표 1–3〉 인터넷 경제 측정 방법론 정립 기본 방향

인터넷 경제 측정 이론 및 방법론 체계 정립

세부 연구 과제	주요 내용
1. 인터넷 경제, 플랫폼 경제 문헌 연구	인터넷 경제의 개념, 포괄 범위 등에 대한 문헌연구 및 해외 사례 연구
2. 네덜란드 통계청의 인터넷 경제 측정 방법론 적용 연구	네덜란드 통계청의 빅데이터 프로젝트 기법 중 하나인 인터넷 경제 측정방법을 우리나라 데이터에 적용하고 우리나라 현실에 맞는 대안 및 개선방안 제시
3. 플랫폼 경제 측정방법론 연구	4차 산업혁명의 기반이 되는 플랫폼 기업(비즈니스)의 경제규모에 대한 측정방법론 제시

○ 우리나라 상황에 적합한 인터넷 경제 측정방법론 제시

– 인터넷 경제/플랫폼 경제 측정 방법론을 구체화하기 위한 전략 체계 제시

〈표 1－4〉 인터넷 경제 측정 연구 수행 전략

기본방향	1 인터넷 경제 측정방법론 개선 방안 도출 2 플랫폼 경제 측정방법론 제시
세부방향	1-1. 인터넷 경제 측정 문헌 및 해외 사례 연구 1-2. 네덜란드 통계청 측정 방법론 적용 연구 1-3. 우리나라 현실에 부합하는 측정 개선방안 연구 2-1. 플랫폼 경제 측정을 위한 해외사례 연구 2-2. 플랫폼 경제 측정방법론 연구

3. 연구 추진 경과

○ 본 연구의 수행을 위해 KISDI 연구진, 아주대 김성환 교수, 통계청 담당자간의 유기적인 연계와 문헌연구, 각종 통계 DB 분석을 통해 과제 수행

－ 세부 연구 방향별 KISDI의 전문 연구진의 역할을 배분해 연구 수행

－ 인터넷 경제(플랫폼 경제)에 대한 측정 방법론 부문은 외부 전문가(아주대 김성환 교수)와의 협력을 통해 수행

○ 특히, 네덜란드 통계청 현지 방문을 통해 네덜란드 통계청, Dataprovider 담당자들과 실제 네덜란드 인터넷 경제 측정 방법론 논의 및 데이터 구매 진행

－ 사업체 분류, 데이터 매칭, Dataprovider의 인터넷 경제 측정을 위한 크롤링 방법론, 데이터 전처리, 통계청 데이터와 매칭 등 데이터 처리 방법 논의

〈표 1－5〉 연구 추진 경과

구분	월일	내용	참석자
전문가 세미나	4. 6	인터넷경제 측정기법 검토	KISDI 연구진, 아주대 김성환 교수
착수보고	5. 24	착수보고	통계청, 연구진
네덜란드 출장	6. 17~6. 22	(네덜란드 통계청) 사업체 분류, 데이터 매칭, 인터넷 경제 측정 방법론 (Dataprovider) 크롤링 방법론, 데이터 전처리, 통계청 데이터와 매칭 등 데이터 처리 방법, 데이터 구매 협의	KISDI 정용찬, 정현준

구분	월일	내용	참석자
중간 보고	8. 23	과제 추진 중간 보고통계승인제도, 통계품질진단	통계청, 연구진
과제 협의	매월	과제 진행 상황 협의(9. 27, 10. 18, 11. 12, 12. 6)	통계청, 연구진
데이터 매칭 통계산출	7. 23~12.	Dataprovider 데이터 구매, 데이터 검증, 통계청 및 공정위, KED 자료와 매칭, 인터넷 경제 규모 산출	KISDI 연구진
최종 보고	12. 11	과제 최종 보고	통계청, 연구진

제 2 장 인터넷 경제 관련 기존문헌 및 해외사례 연구

제 1 절 인터넷 경제 개념과 범위에 대한 문헌 연구

1. 인터넷(또는 디지털) 경제의 개념

가. 국제기구인 OECD(경제협력개발기구)의 인터넷 경제 개념

○ OECD의 인터넷(또는 디지털) 경제의 변천사

- 2010년까지 OECD는 짝수년도에는 정보기술 아웃룩(Information Technology Outlook, 이하 IT아웃룩)과 홀수년도에는 인프라, 요금, 서비스 등 통신정책을 다루고 있는 커뮤니케이션 아웃룩(Communications Outlook)을 발표
- 2012년에 인터넷 경제가 생산성과 경제성장에 미치는 역할의 이해를 돕기 위한 인터넷 경제 아웃룩(Internet Economy Outlook)을 발간
- 2013년 12월부터 디지털경제정책위원회가 운영되면서 2015년[1] 인터넷경제 아웃룩(Internet Economy Outlook) 및 커뮤니케이션 아웃룩(Communications Outlook)을 통합하여 '디지털경제 아웃룩'(Digital Economy Outlook)으로 단일화

○ OECD의 인터넷 경제의 개념 변화

- OECD(2008)는 인터넷 경제를 인터넷과 정보통신기술에 기반한 모든 경제·사회·문화적 활동으로 정의
- OECD(2012)에서는 인터넷을 인터넷프로토콜을 통해 연결성을 제공하는 물리적 네트워크로 인터넷 경제는 인터넷이 제공하는 연결성(connectivity)의

1) 2013년 12월 정보통신 분야 전문위원회인 정보통신정책위원회(Committee for Information, Computer and Communications Policy, 이하 ICCP)가 '디지털경제정책위원회'(Committee on Digital Economy Policy, 이하 CDEP) 로 변경하고 산하 작업반을 3개로 통폐합함

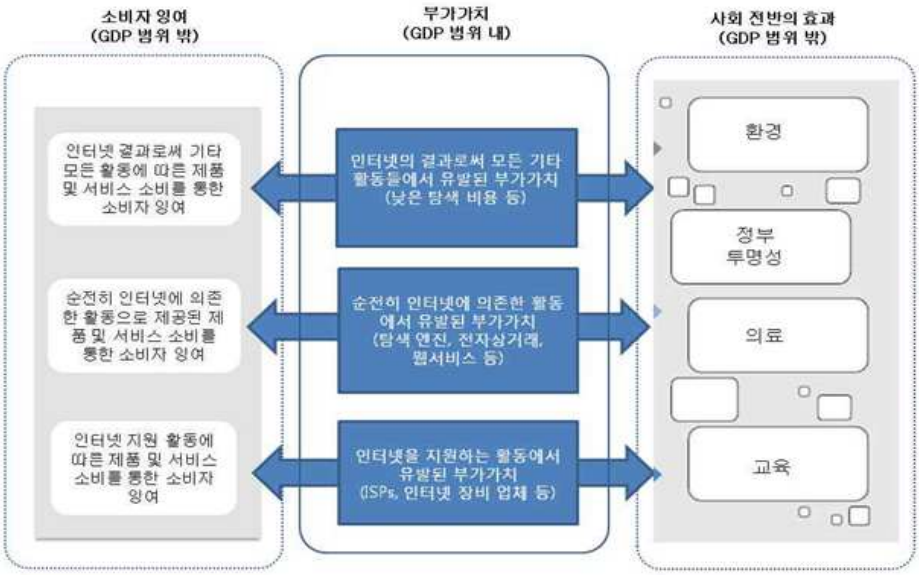
결과 파생되는 모든 이용과 편익을 포함하는 보다 넓은 개념으로 정의

- OECD(2013)에서는 인터넷 경제의 범위를 인터넷과 관련 정보통신기술에 기반한 경제, 사회, 문화 활동의 전 영역으로 정의
- OECD(2015)는 인터넷 경제 대신 디지털 경제라는 용어로 변경, 디지털 경제를 인터넷상에서 전자적 상거래를 통해 재화나 서비스의 거래를 촉진하는 디지털 기술에 기반한 시장으로 정의

○ OECD(2012)의 인터넷 경제의 범위

- 인터넷 경제의 규모를 크게 직접적인 효과(direct impact), 동태적인 효과(dynamic impact), 간접적인 효과(indirect impact)로 구분
- 인터넷 관련 기업은 크게 인터넷 접속 및 서비스 제공업자(ISPs), 데이터 처리 및 웹 호스팅 제공업자, 인터넷 탐색 엔진 및 포털 업체, e-커머스 중개업체, 지불 시스템 업체, 소셜 네트워크 및 콘텐츠 제작 보조 기업 등으로 구분됨

〔그림 2-1〕 OECD(2012)의 인터넷 경제의 파급효과의 범위 규정



자료: OECD(2012)

– 총체적인 인터넷 관련 기업이 인터넷 경제의 중개적 역할(Internet intermediaries)을 수행함으로써 다양한 산업에 속한 기업 간의 커뮤니케이션과 거래를 촉진시키는 인프라 및 플랫폼 기능 수행

– 인터넷경제의 범위를 GDP 범위 내 부가가치와 소비자 잉여, 사회 전반의 효과로 규정함

나. 글로벌 컨설팅 기관의 인터넷 경제 개념

○ Boston Consulting Group(2010)은 인터넷 경제의 파급 효과 분석을 위한 범위를 GDP내에서 파악 가능한 범위로 한정

– GDP 범위 내에서는 파악되지 않지만 소비자 및 기업의 사업 수행에 영향을 미치는 추가적인 파급 효과가 존재한다고 강조

〈표 2－1〉 Boston Consulting Group(2010)의 인터넷 경제 범위

인터넷 경제의 범위	주요 내용
GDP 범위 내 부가가치	• 소비자, 정부, 기업, 순수출 측면에서 파악된 인터넷과 관련된 지출 규모
소비자 및 기업 파급효과	• GDP에 측정되지 않는 B2B 전자상거래, 온라인 광고, 소비자 후생 및 혜택
생산성 파급효과	• 제조업 전자 조달을 통한 생산성 향성, 도소매 거래에 있어서 인터넷 판매를 통한 생산성 향상
사회적 파급효과	• UGC, 소셜 네트워킹, 사기 및 저작권 침해 행위 등을 포함하는 광범위한 사회적인 영향

자료: Boston Consulting Group(2010)

○ McKinsey(2011)의 경우 명확히 인터넷 경제를 정의하기 보다는 인터넷과 관련된 활동들을 인터넷 서비스와 네트워크의 창조 및 사용과 관련된 활동이라고 포괄적으로 규정

– 인터넷 지출 및 소비 측면에서 파악되는 4가지 유형의 활동으로 구분

※ 1) 웹을 보조적으로 사용하는 웹활동, 2) IP에 기초하거나 IP 통신에 연결된 통신 활동, 3) 웹에 연결된 소프트웨어 및 서비스, 4) 하드웨어 제조업자 및 웹 관련된 도구의 유지 업자들의 활동

– 공급 측면에서는 크게 사업자를 통신사업자, 소프트웨어 및 서비스 사업자, 하드웨어 제조 및 유지·보수업자로 구분

〈표 2－2〉 McKinsey(2011)의 인터넷 경제의 범위

인터넷 지출 및 소비 측면	공급 측면
• 인터넷을 활용한 서비스 (예: e-commerce, VoD와 같은 인터넷에서 거래된 콘텐츠, 기타 인터넷 사용(관리 및 gambling), 온라인 광고) • IP에 기반하거나 연결된 통신서비스(예: 브로드밴드, ISP) • 웹에 연결된 소프트웨어 및 서비스(예: IT컨설팅, SW 개발) • 웹관련 하드웨어 제조업자 및 유지업자(예: 컴퓨터, 스마트폰, 하드웨어 장비, 인터넷 서버)	• 통신(예: ISPs, 기존의 스위치 기반 음성 및 GSM 음성 제외 • 소프트웨어 및 서비스(예: IT컨설팅 또는 SW개발) • 하드웨어 제조 및 유지보수(예: 컴퓨터 또는 스마트폰)

자료: McKinsey(2011)

○ 플랫폼 경제란 플랫폼에 의해 실현된 모든 경제·사회적 활동을 의미

〈표 2－3〉 인터넷 경제, 디지털 경제, 플랫폼 경제의 개념 비교

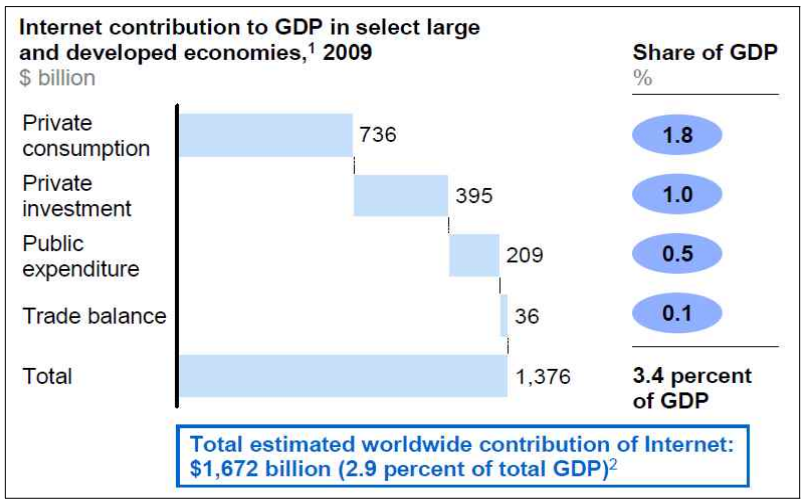
인터넷 경제	○ 인터넷과 관련된 정보통신기술에 기반한 경제, 사회, 문화활동의 전 영역으로 정의
디지털 경제	○ 인터넷상에서 전자적 상거래를 통해 재화나 서비스가 거래되는 촉진하는 디지털 기술에 기반한 시장으로 정의
플랫폼 경제	○ 양면 시장, 온라인 상거래, 공유 경제 등 플랫폼에 의해 실현된 모든 경제·사회적 활동으로 정의

2. 인터넷(또는 디지털) 경제 시장 구모

○ 인터넷 또는 디지털 경제 시장 규모는 인터넷 또는 그 유사 활동들이 GDP에 기여하는 정도로 산출

– McKinsey(2011)는 2009년 기준 세계 인터넷의 기여도 규모가 1조 6,720억 달러이며, 세계 GDP의 2.9%를 차지할 것으로 추정

〔그림 2－2〕 McKinsey(2011)의 인터넷 경제의 규모

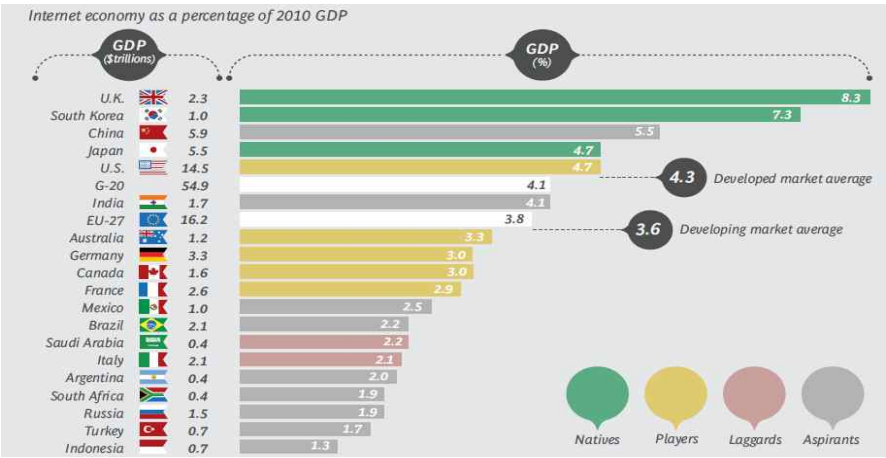


자료: McKinsey(2011)

○ Boston Consulting Group(2012)은 2010년 기준 G20 국가의 인터넷 경제 규모가 약 55조 달러에 이르며, 세계 GDP의 4.1%를 차지할 것으로 추정

※ GDP 대비 인터넷 경제 비중이 높은 국가는 영국이 8.3%로 1위, 한국은 7.3%로 2위, 그 외 중국, 일본, 미국 순임

〔그림 2－3〕 BCG(2012)의 인터넷 경제의 규모



자료: BCG(2012)

○ OECD(2017)의 Digital Economy Outlook에서는 디지털 경제 시장 규모를 전체 산업내 ICT 부가가치 비중으로 산출

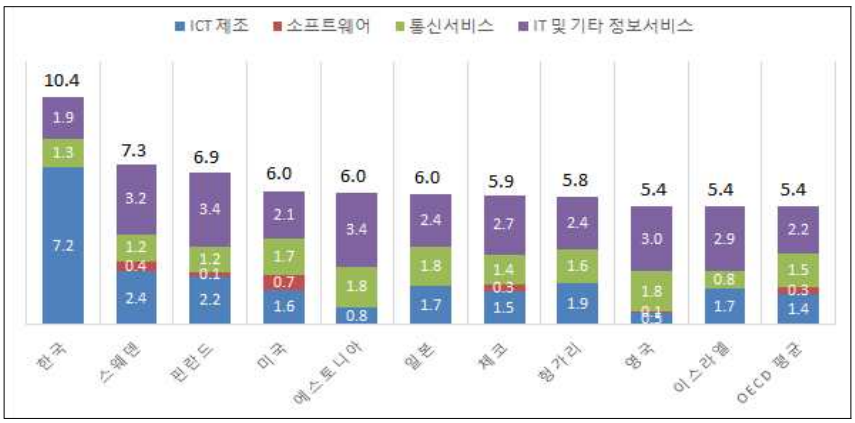
– 2015년 기준 전체 산업내 ICT 부가가치의 OECD 평균은 5.4%임

– 한국의 2015년 기준 ICT 부가가치 비중은 10.4%로 OECD 국가 중 1위를 기록했으며, ICT 제조업 비중이 7.2%로 매우 높고 소프트웨어나 서비스 비중은 매우 낮음

– 반면, 한국을 제외한 OECD 주요국들은 ICT제조업 비중은 상대적으로 낮고 소프트웨어나 서비스 비중이 높은 것으로 나타남

〔그림 2–4〕 OECD 주요국의 ICT 부가가치 비중(2015)

(단위: %)



주: 1) 상위 10개국 기준

2) ICT 산업 분류(ISIC rev.4 기준): 26 컴퓨터 전자 및 광학 기기, 582 소프트웨어 퍼블리싱 61 통신, 62-63 IT 및 기타 정보서비스

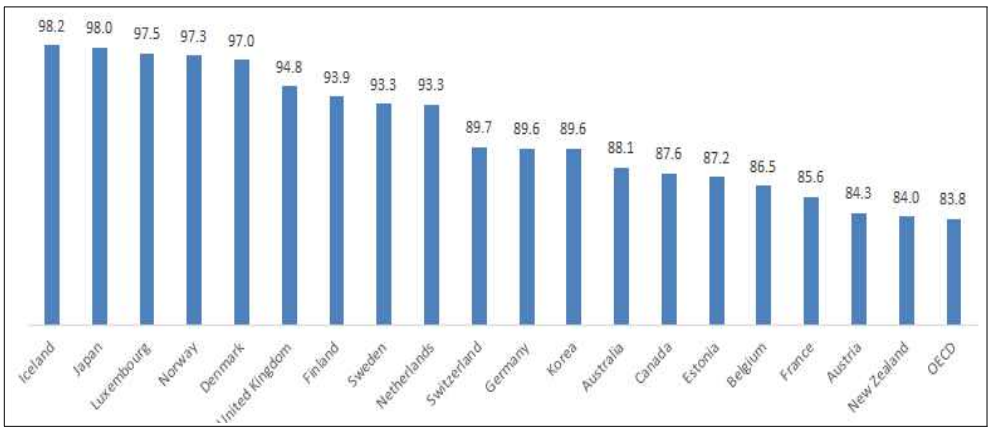
자료: OECD Digital Economy Outlook 2017

○ OECD 주요국의 디지털 경제 활동 현황

– 전체 인구대비 인터넷 사용률은 2016년 기준 OECD 평균이 83.8%이며, 한국의 인터넷 사용률은 89.6%로 OECD 국가 중 12위 기록

〔그림 2－5〕 OECD 주요국의 인터넷 사용률(2016)

(단위: %)



자료: OECD Digital Economy Outlook 2017

○ OECD 주요국의 웹사이트 및 홈페이지 사용 현황은 2016년 기준 평균 77%이며 2010년에 비해 7.4%p 상승

－ 한국 기업은 2016년 기준 평균 60.1%가 웹사이트 또는 홈페이지를 사용해 OECD 국가 중 하위권이며, 특히 50인 미만의 소기업들의 홈페이지 사용률은 57.4%로 매우 저조

〈표 2－4〉 OECD 주요국 기업들의 웹사이트 or 홈페이지 사용 현황(2010, 2016)

(단위: %)

구분	2010년 전체 기업	2016년 전체 기업	16-10년 상승 비율	10-49인 기업	50-249인 기업	250인 이상 기업
Finland	87.3	95.3	8.0	94.4	99.7	100.0
Denmark	87.8	93.3	5.5	92.6	96.4	98.2
Switzerland	90.1	91.7	1.7	90.5	95.7	99.1
Sweden	88.7	89.7	1.1	88.7	95.0	97.4
Germany	81.4	88.9	7.5	87.7	93.5	96.9
Netherlands	81.5	88.7	7.2	87.4	93.5	96.3
Japan	91.5	88.6	－2.9	－	85.9	94.3
Austria	80.2	88.1	7.9	86.7	95.3	98.6
Iceland	77.2	84.7	7.5	72.9	89.7	92.2
United Kingdom	75.7	83.0	7.3	80.6	94.2	97.2

구분	2010년 전체 기업	2016년 전체 기업	16-10년 상승 비율	10-49인 기업	50-249인 기업	250인 이상 기업
Slovenia	72.8	82.6	9.8	79.8	94.1	100.0
New Zealand	68.8	82.2	13.4	80.5	90.4	97.8
Czech Republic	73.6	82.1	8.5	79.2	92.8	95.1
Luxembourg	70.2	81.3	11.1	78.7	90.7	98.3
Belgium	78.5	81.0	2.5	78.9	90.4	96.3
Norway	78.4	80.4	1.9	78.2	92.5	96.3
Slovak Republic	74.1	78.3	4.2	76.8	83.0	85.6
Estonia	70.0	77.9	7.9	75.1	89.6	95.8
Canada	69.7	77.5	7.8	73.7	88.8	91.5
Lithuania	65.0	77.0	12.0	73.4	90.6	98.8
OECD	69.6	77.0	7.4	73.9	88.2	94.0
대한민국	60.1	59.7	0.4	57.4	73.4	87.9

자료: OECD Digital Economy Outlook 2017

제2절 인터넷 경제 측정 방법론에 대한 사례 연구

1. 인터넷 경제 측정 방법론 개요

○ 인터넷 경제 규모 측정을 위한 주요 방법론은 크게 산출 방식에 따른 방법론과 산출 범위에 따른 방법론으로 구분

○ 산출 방식에 따른 방법론은 총 지출 기반의 접근 방식과 총 소득 기반의 접근 방식으로 구분

– 총 지출 기반의 접근 방식은 인터넷 경제의 범위를 규정하고 난 후, 규정한 범위에 해당하는 경제주체의 지출을 합하거나 규정한 범위에 해당하는 상품 및 서비스에 대한 지출을 합산하는 방법론임

– 총 소득 기반의 접근 방식은 인터넷에 기반한 사업체 및 사업체 종사자의 소득을 합산하는 방식으로 인터넷 경제 규모를 측정하는 방법론임

○ 산출 범위에 따른 구분은 직접적 효과, 동태적 효과, 간접적 효과로 구분

– 직접적 효과 방식은 전통적인 방식의 GDP 산출에 사용되는 공식적이고 개방된 데이터를 활용하여 전체 GDP 중 인터넷 경제에 해당하는 부분을 구분하

는 형식의 방법론임

※ 인터넷 경제 관련 업종을 정의해 해당 업종의 부가가치를 합산하는 방식

– 동태적 효과 방식은 인터넷이 총생산에 기여하는 직접적인 부분과 더불어 인터넷으로 인한 생산성 및 수익성 향상 등 인터넷으로 인한 동태적인 경제 기여분을 포함하여 인터넷 경제의 규모를 추정하는 방법론임

※ 일반적으로, 인터넷으로 인한 경제성장 기여율을 다양한 방식으로 추정하여 전통적인 방식으로 산출된 GDP에 적용

– 간접적 효과 방식은 인터넷의 도입으로 인한 소비자 잉여 및 사회후생의 증가분 등 전통적인 GDP 산출방식으로 평가되지 않는 인터넷의 경제적 가치를 인터넷 경제 규모에 포함하여 추정하는 방법론임

※ 삶의 질적인 측면과 함께 디지털 시대에 등장한 공유경제 등 새로운 거래 방식을 포함하는 'Beyond GDP'와 연관

〈표 2–5〉 산출 방식에 따른 인터넷 경제 규모 측정 방법론

주요 방법론		특징	한계점	선행연구
산출방식에 따른 구분	총지출 기반 방법론	• 인터넷 경제의 범위를 규정하고 난 후, 규정한 범위에 해당하는 경제주체의 지출을 합하거나 규정한 범위에 해당하는 상품 및 서비스에 대한 지출을 합산하는 방법론	• 인터넷 경제에 해당하는 경제주체의 지출 전체가 인터넷 경제에 포함됨으로 인한 과대추정의 문제가 존재	• Boston Consulting Group(2010) • McKinsey(2011) • Shin et al.(2012)
	총소득 기반 방법론	• 인터넷에 기반한 사업체 및 사업체 종사자의 소득을 합산하는 방식으로 인터넷 경제 규모를 측정하는 방법론	• 인터넷에 기반하지 않은 사업체나 종사자의 소득 중 일부도 인터넷 관련 활동을 통한 소득일 수 있으나 이를 누락함으로써 인터넷 경제규모가 과소 추정될 수 있는 문제가 존재	• Hamilton Consultants(2009)
산출범위에 따른 구분	직접적 효과	• 전통적인 방식의 GDP 산출에 사용되는 공식적이고 개방된 데이터를 활용하여 전체 GDP 중 인터넷 경제에 해당하는 부분을 구분하는 형식의 방법론 • 일반적으로 해당 업종의 부가가치를 합산하는 방식	• 인터넷 경제의 범위를 산업이나 업종 단위에서 규정함으로 인하여 인터넷 경제의 유동적인 속성을 반영하지 못함	• Shin et al.(2012) • McKinsey(2011) • Deloitte(2011) • Boston Consulting Group(2010) • Hamilton Con-sultants(2009)

주요 방법론		특징	한계점	선행연구
산출범위에 따른 구분	동태적 효과	• 인터넷이 총생산에 기여하는 직접적인 부분과 더불어 인터넷으로 인한 생산성 및 수익성 향상 등 인터넷으로 인한 동태적인 경제 기여분을 포함하여 인터넷 경제의 규모를 추정하는 방법론 • 인터넷으로 인한 경제성장 기여율을 다양한 방식으로 추정하여 전통적인 방식으로 산출된 GDP에 적용	• 연구자에 따라 인터넷으로 인한 경제성장 기여율의 추정치가 매우 크게 차이가 나 정책입안 등 실용적인 활용성이 떨어짐	• Forman and van Zeebroeck (2010) • Polder et al.(2009) • Bertschek et al. (2011)
	간접적 효과	• 인터넷의 도입으로 인한 소비자 잉여 및 사회후생의 증가분 등 전통적인 GDP 산출방식으로 평가되지 않는 인터넷의 경제적 가치를 인터넷 경제 규모에 포함하여 추정하는 방법론 • 'Beyond GDP'와 연관	• 인터넷 도입으로 인한 소비자 잉여 및 사회후생을 측정하는 방법론은 아직 초기단계에 있어, 이를 활용한 인터넷 경제 규모 추정은 이론적인 수준에 머물고 있음	• Grochen et al. (2017), • Ahmad and Schereyer(2016)

2. 인터넷 경제 측정 방법론 관련 주요 사례 분석

가. 산출 방식에 따른 방법론

○ Dilloite(2011)[2]의 추정 방법론

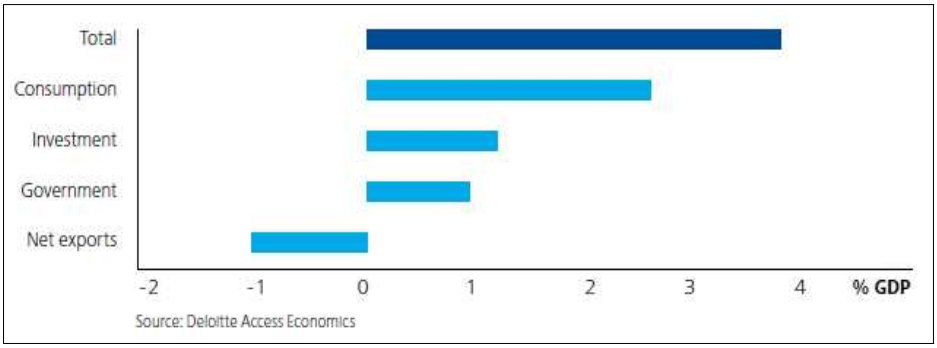
– 인터넷 경제의 직접 기여 규모, 즉 GDP 대비 인터넷 경제 비중을 세 가지 방식으로 추정했는데 세 가지 모두 이론적으로 동일하게 부가가치의 총합으로 산출

– 순생산 기반 접근: 7개의 업종을 인터넷 경제로 정의하고 여기에 포함된 사업체의 총매출에서 총투입을 제한 부가가치창출액으로 추정

– 총지출 기반 접근: 소비지출, 투자지출, 정부지출, 순수출의 특정 부문을 인터넷 경제에 대한 지출로 분류하여 이들의 총합으로 추정

2) Dilloite(2011). "The Connected Continent I, II: How the Internet is Transforming the Australian Economy".

〔그림 2-6〕 총지출 기반 접근방식으로 추정된 호주의 GDP 대비 부문별 인터넷 경제 비중



– 총소득 기반 접근: 7개의 업종을 인터넷 경제로 정의하고 여기에 포함된 기업 및 종사자의 총소득을 인터넷 경제의 규모로 정의

〔그림 2-7〕 총소득 기반 접근방식으로 추정된 호주 인터넷 경제의 업종별 종사자 수 및 부가가치 총액

Segment	Internet employees no.	Value-added $b
ISPs, web search portals and data processing	15,000	3.7
Hardware	4,000	0.6
IT software and consulting	40,000	4.5
Online information services	4,000	0.6
Advertising and enterprise sites	75,000	5.6
Government	2,000	2.3
E-commerce	50,000	4.9
Total	**190,000**	**22.3**

Source: Deloitte Access Economics, ABS, IBISWorld, eBay

– 각 방식은 데이터 접근성 및 측정 용이성의 측면에서 한계점을 내재
– 순생산/총지출 기반 접근: 많은 산업분야에서 인터넷으로 인한 부가가치를 분리하여 측정하는 것이 매우 어려움
– 총소득 기반 접근: 인터넷 경제에 포함된 기업 및 종사자의 소득 전체를 인

터넷 경제로 규정할 경우 지나친 과대추정의 가능성이 존재

○ 신일순 외(2012)[3] 인터넷 경제 측정 방법론

– 총지출 기반 접근 방법을 통해 대한민국 GDP 중 인터넷 경제의 비중 및 경제성장 기여율 추정

– 인터넷 관련 지출을 “활동”, “접근”, “디바이스” 지출로 분류

– ① “활동” 지출: 인터넷과 관련된 모든 경제활동에 대한 지출

– ② “접근” 지출: ISP에 대한 지출 등 인터넷 접근을 위한 지출

– ③ “디바이스” 지출: 인터넷과 관련된 하드웨어 및 소프트웨어에 대한 지출

〈표 2−6〉 지출 분류별 세부지출항목

분류	포함되는 세부지출항목
소비지출	• B2C 전자상거래 지출 • C2C 전자상거래 지출 • 온라인 콘텐츠 대상 지출 • 온라인뱅킹 수수료 • 인터넷 서비스 비용(ISP) • 관련 장비 대상 지출
투자지출	• 통신업체 투자지출 • 이외 민간기업 통신관련 투자지출
정부지출	• B2G 전자상거래 지출 • ICT 관련 정부지출
수입 및 수출 관련 지출	• 하드웨어 및 소프트웨어 순수출액

– 각 지출 분류는 다시 개인(소비), 정부, 기업(투자), 그리고 해외(수출) 부문으로 분류

– 다른 연구들이 전자상거래 등 소비 지출 또는 개인+기업에 대한 지출만 연구한 경우가 많았으나 본 연구의 범위는 개인, 정부, 기업, 수출 등을 모두

3) Ilsoon Shin, Jungsoo Park, Seonghoon Jeon(2012). “The Size, Effects, and Efficiency of Internet Economy in Korea”, *International Telecommunications Policy Review,* Vol. 19, No. 2, 2012

포함해 포괄적인 인터넷 경제 규모를 측정함

– 각 세부항목별 활용 데이터는 2005~2010년 자료를 활용하였으며, 산업연관표(IO Table) 자료는 그 기간 내 수집 가능한 2005년과 2008년 데이터를 사용하였음

– 본 연구를 통해 측정된 한국의 인터넷 경제 규모는 GDP의 5.94%로 나타남

〈표 2－7〉 인터넷 관련 지출의 항목 별 활용데이터

GDP Component	Category	Detail	Data
Consumption	Activity	E-Commerce(Including C2C E-Commerce)	Statistics Korea (The Statistics on E-commerce)
		Contents	Ministry of Culture, Sports and Tourism (The Statistics on Cultural Industries)
		Online Banking Fees	Financial Supervisory Service (E-Commerce Handling Performance)
	Access	Expenditure on Internet Access	Input-Output Tables (2005, 2008)
	Device	Computer, HW, SW	Input-Output Tables (2005, 2008)
Government Expenditure	Activity	Government E-Commerce (Purchasing of Goods and Services)	The Statistics on E-commerce (Statistics Korea)
	Access	Expenditure on Internet Access	Input-Output Tables (2005, 2008)
	Device	Computer, HW, SW	Input-Output Tables (2005, 2008)
Investment	Activity	Investment of Telecommunications Companies	Statistical Yearbook of Broadcasting and Communications Industry
	Access	Expenditure on Internet Access	Input-Output Tables (2005, 2008)
	Device	Computer, HW, SW	Input-Output Tables (2005, 2008)
Net Export	Device	Computer, HW, SW	Input-Output Tables (2005, 2008)

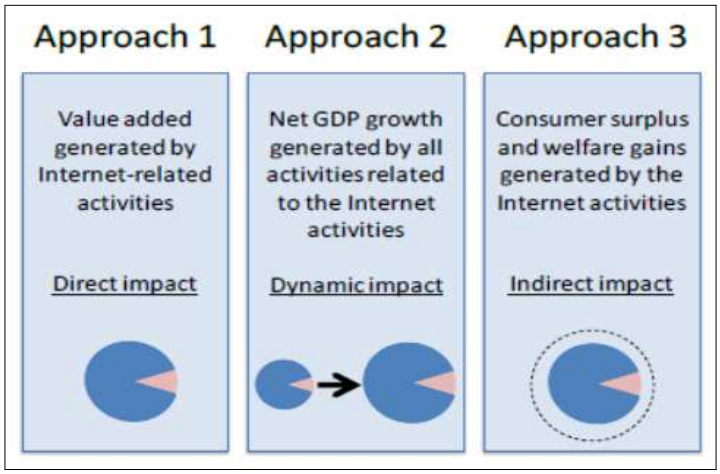
자료: 신일순 외(2012)

나. 산출 범위에 따른 방법론

○ OECD[4]의 인터넷 경제 측정 방법

– OECD는 “OECD Internet Economy Outlook 2012”에서 인터넷 경제의 규모에 대한 접근 방식을 직접적 효과(direct impact), 동태적 효과(dynamic impact), 간접적 효과(indirect impact)로 구분

〔그림 2–8〕 인터넷 경제규모 측정의 세 가지 접근 방식



– 직접적 효과(Direct impact) 접근방식은 가장 보수적인 방식으로 인터넷 경제를 규정하는 방식이며 공식적이고 비교적 접근이 쉬운 데이터(통상적인 GDP 측정에 활용되는 데이터)를 통해 측정이 가능

– 동태적 효과(Dynamic impact) 접근방식은, 종국적으로 인터넷은 모든 산업과 경제활동에 영향을 미친다는 가정 하에 각 경제활동에서 인터넷이 미치는 경제적 영향(생산성, 수익성 등)을 동태적인 관점에서 측정하고자 하는 방식

– 간접적 효과(indirect impact) 접근방식은 인터넷의 간접적인 영향, 예컨데, 인터넷으로 인한 소비자 잉여, 사회적 편익 등 비수량적(non-quantitative) 영

4) OECD(2013). “Measuring the Internet Economy: A Contribution to the Research Agenda”.

향까지 측정하고자 하는 방식

○ 인터넷경제의 거시경제학적 규모 측정을 위해선 미시경제학적(Firm-level)분석이 기초가 되어야 함을 강조(BigData 분석의 필요성 대두)

– 인터넷경제의 규모를 측정함에 있어 인터넷 기술을 일반목적기술(GPT)로 간주하는 것이 필요

〈표 2－8〉 OECD(2012)의 인터넷 경제의 범위 및 추정 방법

접근 방법	주요 내용
직접적 효과	– 전체 GDP내에서 인터넷을 지원하거나 인터넷 상에서 발생하는 경제적 활동으로 인한 부가가치를 추정 ① 인터넷의 운용 및 사용과 관련된 활동(브로드밴드 장비 제조 업체, ISPs) ② 온전히 인터넷에 기반한 활동(서치 엔진, 전사상거래 웹서비스)
동태적 효과	– 인터넷의 활용이 GDP성장에 미치는 순효과(net effect)를 추정 – 기업, 산업, 지역, 국가별 자료를 이용한 실증분석을 통해 인터넷의 효과를 추정 – 전체적인 관계를 보기 때문에 세부적인 관계 및 타산업 상쇄효과에 따른 인터넷 전체 효과 미파악
간접적 효과	– GDP에 포함되지 않는 소비자 후생과 같은 간접 효과를 추정 – 인터넷으로 인한 복지 향상(비금전적인 거래관계, 사회자본 형성, 환경, 교육, 의료, 정부투명성에 미치는 영향)

3. 인터넷 경제와 플랫폼 경제 간의 연관성

○ 플랫폼 경제는 일반적으로 “디지털” 플랫폼 경제를 일컬으며, 플랫폼에 의해 실현된 모든 경제·사회적 활동을 포함함

– 디지털 플랫폼은 보통 거래중계 플랫폼(transaction platform)과 혁신 플랫폼(innovation platform)으로 구분되며, 이들은 모두 인터넷 경제의 일부분으로 취급됨

– 다만, 플랫폼 경제에 속한 영역은 “양면경제”의 속성을 가지고 있어 전통적인 방식으로 측정할 경우 그 규모가 과소 추정될 여지가 있음

○ Zervas et al.(2017)에서 Airbnb가 텍사스 오스틴 시의 호텔 산업에 미친 영향을 추정한 결과

- 기존 호텔들은 Airbnb 등장으로 인해 평균 8~10%에 이르는 수입이 감소했음
- 이와 같은 수입 감소는 Airbnb의 매출증가분 외에도 호텔이용자들의 편익 증가, 숙박비용의 절감으로 인한 기타 지출 증가 등의 효과로 파급되므로 해당 연구에서 플랫폼 경제(Airbnb의 경제적 효과)의 규모를 단순히 Airbnb의 매출만으로 규정할 수 없음
- 또한, Botsman(2012)에서 강조한 바와 같이, 플랫폼 경제를 아우르는 이른바 “공유경제(sharing economy)”의 비화폐적 소비자후생 향상은 정보통신기술이 진보함에 따라 무시할 수 없는 수준에 이르고 있음

○ 이와 같은 사례로 볼 때 인터넷 경제의 규모를 측정함에 있어서 기존의 전통적인 방식으로는 올바른 측정이 이루어지지 않을 가능성이 큼

○ 따라서 인터넷 경제를 구성하는 단위를 경제주체별 인터넷 활용 방식 등으로 새롭게 분류하고 기존 통계에서 노출되지 않는 인터넷의 기여분을 측정할 수 있는 새로운 측정방식을 고민해야 함

제 3 장 인터넷 경제 측정 방법론 연구

제 1 절 네덜란드 통계청의 인터넷 경제 측정 방법론 적용 연구

1. 네덜란드 통계청(BOD)의 인터넷 경제 측정 방법론 연구

○ (연구 목적) 인터넷 경제의 중요성에 대한 이해도 증진을 목표로 다음과 같은 세부 목표를 설정, 연구 진행

- 인터넷 경제의 정의를 실체 확인과 빅데이터 활용 가능성이라는 실용적인 관점에서 접근
- 네덜란드 인터넷 경제의 규모와 전체 경제에서 차지하는 비중과 중요성을 탐색
- 통계 생산을 위한 새로운 측정 방법론의 가능성을 모색
- 현재 적용하고 있는 통계의 개념, 표준 분류와의 차이점을 탐구

○ (분류 체계) 인터넷과의 연관성 관점에서 사업체를 분류

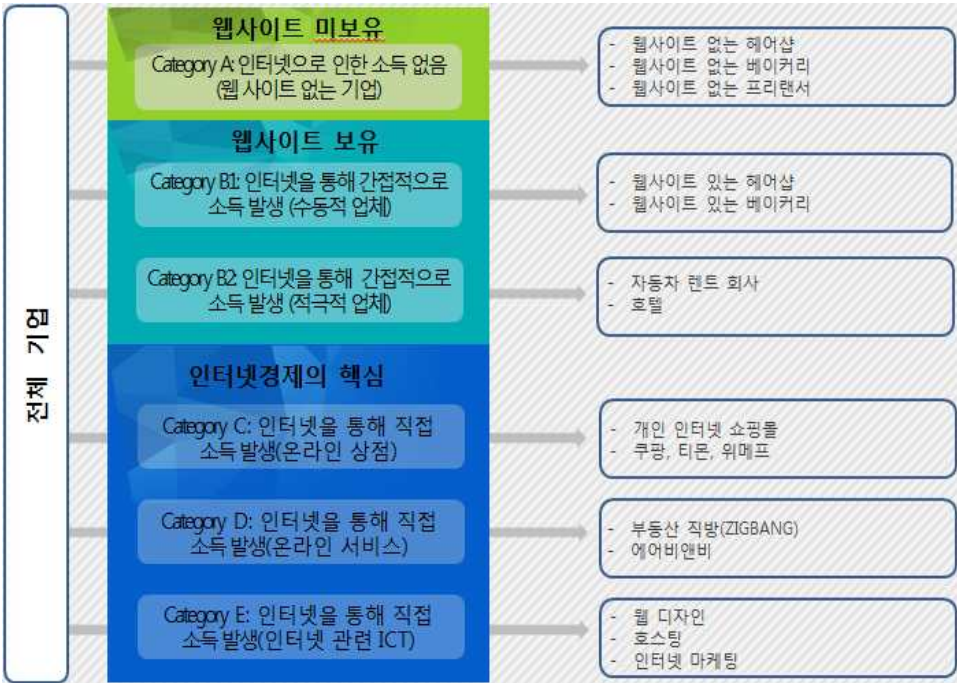
- 웹데이터 크롤링 전문회사인 Dataprovider*가 네덜란드 정부와 산업계, 학계, 구글**과 협력하여 도출한 분류체계 활용

* 네덜란드 기업으로 40개국 1억 개 이상 웹사이트에 대한 정보를 크롤링한 데이터셋을 외부에 판매

** 프로젝트 자금 지원, 관련 연구 공동수행

- 기업체가 인터넷을 어떻게 사용하는지에 따라 웹사이트 미보유(A), 웹사이트 보유(B), 온라인 상점(C), 온라인 서비스(D), 인터넷 관련 ICT(E) 기업체로 분류
- 분류 B~E를 인터넷 경제 활동 기업으로 정의하고, 이중 C~E를 인터넷 경제의 핵심(core)으로 분류

〔그림 3－1〕 인터넷 활용에 따른 기업 분류



출처: CBS(2016), Measuring the internet economy in The Netherlands: a big data analysis 재구성

○ (활용 데이터) 민간회사의 웹사이트 크롤링 데이터와 네덜란드 통계청의 경제 데이터를 활용

– 웹사이트 크롤링 데이터는 사이트명, 업체명, 사업자등록번호(CoC-chamber of commerce), 접속 횟수 등으로 구성되어 있으며 이 외에 주문, 구매, 쇼핑카트, 예약 등의 이용자 데이터베이스로 구성

– 네덜란드 통계청의 경제 데이터는 기업 규모(종업원 수), 분야, 업력 등으로 구성된 기업등록부(General Business Register, GBR)와 재무 정보로 구성된 생산통계(Production Statistics)를 활용

– 이 외에 부가가치세 데이터, 매출, 지역 통계, 행정자료, 소매조사(Retail Survey), ICT 활용 조사 등을 활용

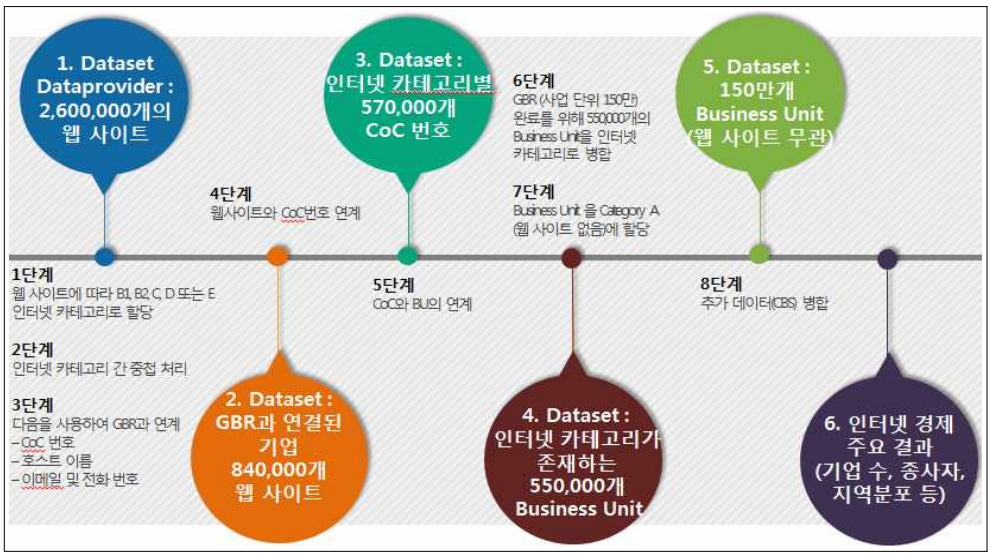
〈표 3－1〉 인터넷 경제 측정을 위한 데이터

구분	내용
인터넷 데이터 (Internet data)	－ 네덜란드 웹사이트 리스트: 업체명, 사업자등록번호(CoC-chamber of commerce-number) 접속 횟수 등 － Call To Action(CTA) database: 주문, 구매, 쇼핑 카트, 예약, 구독, 등록 등 웹사이트 이용자 데이터
경제 데이터 (Economic data)	－ 기업등록부(General Business Register, GBR): 기업 규모(종업원수), 분야, 업력 － 생산통계(Production Statistics): 재무 정보(50인 이상 기업 전수조사)
부가가치세 데이터(Baseline)	－ 생산가치(Production Value), 부가가치(Value Added) － 취업자 수(Employed Persons) － 전일 등가물(Full-Time Equivalents): 임의의 업무에 투입된 노동력을 전일 종사자 노동자 수로 측정
매출 데이터 (Turnover Statistics)	－ 부가가치세(VAT)를 제외한 1차 및 2차 산업 활동으로부터의 수익
지역 통계 (Regiobase)	－ Local Business Unit(LBU): 우편번호가 포함된 지역차원의 데이터
행정자료	－ Policy Record Administration: 소득세 신고를 통해 수집된 모든 근로자의 고용기록
소매조사 (Retail Survey)	－ 소매분야에 10명 이상의 직원이 있는 기업의 표본
ICT 활용 조사 (ICT use survey)	－ 기업의 자동화와 정보통신기술(ICT)사용에 관한 연간데이터

출처: CBS(2016)

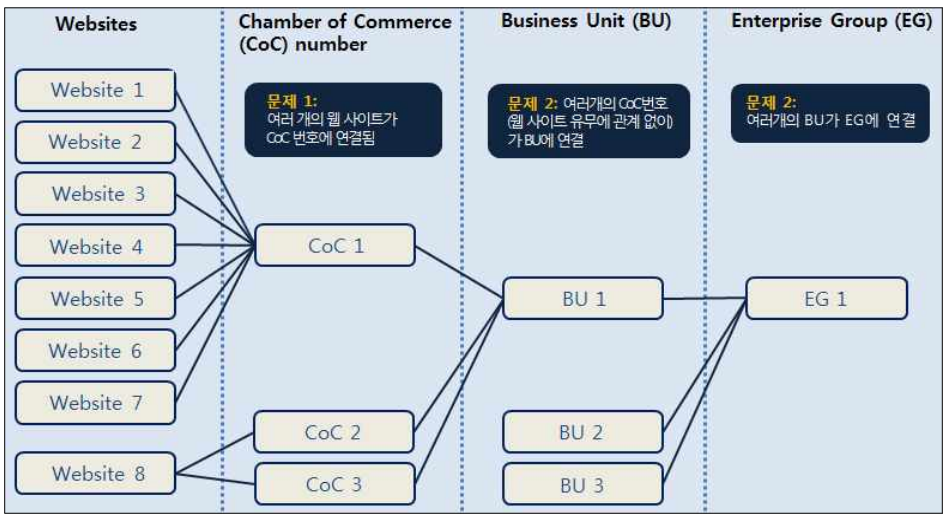
○ (데이터 연계 방법론) 민간회사의 웹사이트 데이터를 네덜란드 통계청의 사업자등록번호(CoC), 기업체(BU), 그룹사(EG) 데이터와 연계

〔그림 3－2〕 데이터 연계 방법론



출처: CBS(2016)

〔그림 3－3〕 웹사이트와 사업체 기업 연계 방법



자료: CBS(2016), p. 25.

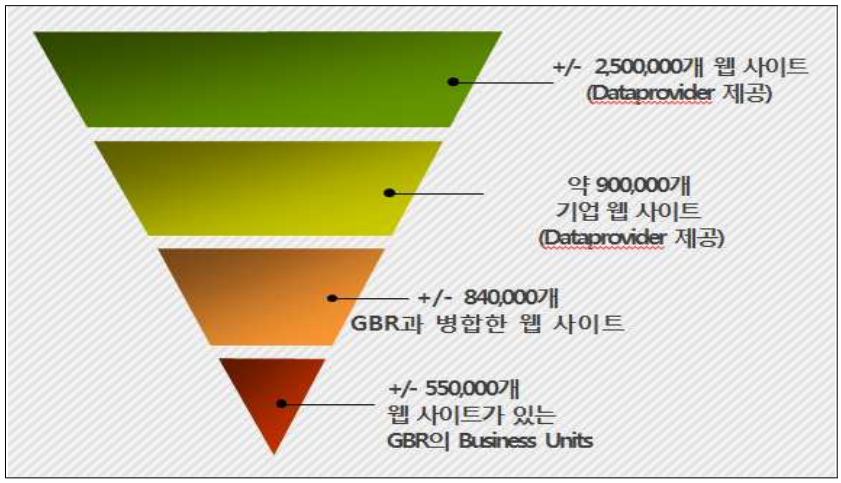
○ (웹 사이트 연계방식) 웹사이트를 사업자등록번호에 연계하여 BU, EG 데이터를 도출

– 여러 개의 웹사이트가 CoC번호에 연결되고, 여러 CoC번호가 단일 BU에 연결되는 문제가 발생하기도 함

– 웹 사이트에서 CoC번호, BU 수준으로 변환하는 것과 관련된 문제에 대응하기 위하여 일련의 규칙을 개발할 필요가 있음

○ (방법론에 따른 웹사이트 병합) Dataprovider에서 제공하는 웹사이트를 연계 방법론에 따라 병합

〔그림 3－4〕 웹사이트 정제 후 최종 기업체 수



자료: CBS(2016), p. 26.

○ 주요 결과는 다음과 같음

– 2015년 기준 36%의 기업이 웹사이트를 보유

– 웹사이트를 보유하지 않은 기업의 83%가 자영업자

– 코어인터넷 기업(온라인 스토어, 온라인서비스, 인터넷 관련 ICT)은 5만개로 종사자는 345,000명(4.4%), 1040억 유로(7.7%) 규모로 주로 건설, 숙박 및 음식, 교통 분야

– 온라인 스토어의 절반이 소매업에 해당(표준산업분류 기준)

– 새로운 카테고리 분류는 5,700개의 기업으로 구성, 26,000개의 일자리를 제공하여 2015년 100억 유로 연간매출 달성

〈표 3－2〉 네덜란드 사례의 주요 분석 결과

구분	내용
홈페이지 운영 여부에 따른 비중 분석 (A or B 그룹)	– 업종별 / 기업규모(종업원수)별 홈페이지 운영 사업체(B 그룹) 수 – 업력별 홈페이지 운영 사업체(B 그룹) 비중 – 인터넷 활용에 따른 그룹별 사업체, 취업자, 매출액 및 부가가치액 비중 – 인터넷 활용에 따른 그룹별 노동생산성 – 지역별 홈페이지 미운영 사업체(A 그룹) 비중
인터넷 활용 정도에 따른 비중 분석 (B or C,D,E 그룹)	– 일정규모(종업원수) 이상의 사업체 중 인터넷 활용 정도에 따른 사업체 (B ~ E 그룹) 비중 – 지역별 / 인터넷 활용 정도에 따른 사업체(B~E 그룹)별 비중 – 업력별 / 인터넷 활용 정도에 따른 사업체(B~E 그룹)별 비중 – 타산업과 코어인터넷기업(C~E 그룹) 간의 사업체 및 취업자 수, 매출액 비교

자료: CBS(2016), p. 28~40. 재구성

제 2 절 우리나라 데이터 적용 분석

1. 데이터

○ (데이터 가용성) 네덜란드 모델은 민간회사의 웹사이트 크롤링 데이터와 네덜란드 통계청의 경제 데이터를 활용한 것으로 네덜란드 사례와 같이 웹사이트 크롤링 데이터를 한국에 적용 가능한지 검토가 필요

– 이를 위해 데이터프로바이더의 한국 데이터의 구조 파악이 선행되어야 하는데 우선, 사업체번호의 활용가능 여부, 표시 언어(영문 혹은 한글), 크롤링 이후 보완작업의 정도 등을 파악

○ (데이터 입수) 본 연구 수행을 위해 통계청의 기업등록부 자료를 입수하고, 데이터프로바이더로부터 웹크롤링 데이터를 구입

– 데이터 수집 및 분석은 기업등록부 보안을 위해 통계청의 통계빅데이터센터에서 수행

○ (데이터프로바이더 데이터) 데이터프로바이더의 한국 데이터는 925,961개의 웹사이트를 크롤링한 데이터로 구성

– 데이터프로바이더 데이터는 6개 섹션 128개 변수로 구성(세부 내용은 부록 참조)

※ 네덜란드 웹사이트 크롤링 데이터는 사업체와 관련한 사이트 정보와 이용자와 관련한 데이터베이스로 구성

○ (데이터프로바이더 웹사이트 및 키변수) 통계청의 경제 데이터와 연계를 위한 key 변수를 파악

– 1차적으로 사업자번호, 법인번호 등의 활용 가능성을 점검하고,

– 2차적으로 확률매칭을 위한 사업체명, 전화번호, 주소 등의 활용 가능성을 점검. 이때, 영문의 한글 전환 등이 필요한지 등을 확인한 결과 데이터 전처리 후 전화번호 매칭이 가능할 것으로 판단함

〈표 3－3〉 주요 분석 결과

구분	명칭	레코드	정확도	커버리지
호스팅 정보	웹사이트주소 (Hostname)	928,218개	100%	-
호스팅 정보	도메인주소	928,218개	100%	-
키변수	사업체명 (company name)	146,331개	90%	16%
키변수	사업자번호 (Business Registry number)	13,625개	93%	1%
키변수	전화번호 (Phone number)	273,920개	97%	30%
키변수	보조 전화번호 (Secondary phone numbers)	153,383개	97%	17%

※ 웹사이트에서 제시하는 레코드 수는 다운로드 시점의 차이로 소폭 차이가 나며, 데이터 정제 과정에서 일부 차이가 발생

○ (데이터프로바이더 변수 특성) 웹사이트의 특성을 결정할 수 있는 분석에 사용되는 주요 변수 및 개별 속성은 다음과 같음

– (사업체특성) 웹사이트 타입(Website type): 레코드 928,195개, 커버리지 80%. 예: blog, forum, business, eCommerce, placeholder, parking, empty, framed or disallowed

– (사업체특성) 범주(Category): 레코드 103,627개, 정확도 66%, 커버리지 11%

– (e 커머스) 온라인상점(Online store): 레코드 928,195개, 정확도 79%, 커버리지 100%

– (e 커머스) 쇼핑카트시스템(Shopping cart system): 레코드 100,170개, 정확도 86%, 커버리지 1%

– (e 커머스) 배달서비스(Delivery services): 레코드 1,839개, 정확도 67%, 커버리지 0%

– (e 커머스) 지불수단(Payment methods): 레코드 3,766개, 정확도 87%, 커버리지 0%

○ (기업등록부 정보) 통계청의 기업등록부는 사업체명, 사업자등록번호, 법인번호, 전화번호, 주소, 매출액, 종사자수 등의 변수로 구성

– (키변수) 기업등록부의 키변수는 사업자등록번호를 이용했고, 필요에 따라 법인번호, 전화번호를 보조적으로 이용

○ 데이터프로바이더 정보를 이용하여 개별 웹사이트를 “인터넷 사용에 따른 업체분류”의 카테고리 A~E를 부여

2. 데이터셋 매칭 전략

○ (매칭) 네덜란드 방법론을 적용하기 위해서는 매칭에 관한 연구 및 실무적인 적용이 필수적임

– 웹사이트→사업체→기업을 연계 매칭하기 위해 통계매칭, 정확매칭 등의 활용 범위를 검토하고, 적절한 매칭 방법론을 적용하여 매칭 시행

〈표 3－4〉 매칭 방법론 분류

매칭 방법론의 분류			내용
자료연계 recode linkage	통계적매칭 statistical matching		모집단이 같으나 두 개의 자료를 연결할 수 있는 변수가 부재한 경우 마이크로데이터 레벨에서 Imputation의 방법으로 매칭을 하거나 집계표 수준에서 통계적으로 매칭하는 방법
	정확매칭 exact matching	결정매칭 deterministic matching	매칭 대상 변수가 일치할 경우 Match 그렇지 않을 경우 Non-match로 분류하는 방법
		확률매칭 probabilistic matching	철자, 스캐닝 오류 등을 고려하여 두 레코드의 매칭확률을 계산하는 기법

자료: 통계청(2018), 통계빅데이터센터 온라인서비스 및 자료연계시스템 개발 ISP 수립 중간보고 자료

○ (데이터 연계 방법론) 네덜란드 방법론을 차용하여 웹사이트 데이터를 한국 통계청 기업등록부와 연계

– 통계청의 기업등록부는 사업자번호와 법인번호를 이용하여 사업체-기업간 관계를 파악이 가능하도록 정리되어 있음

– 웹사이트 정보에서 사업체→기업으로 추적 매칭하고, 매칭된 사업체 혹은 기업 단위에서 조사된 정보를 분석

○ (데이터프로바이더) 데이터프로바이더의 데이터는 다음과 같은 형태로 구성됨

– 데이터셋 매칭을 위한 키변수로 사업자번호와 전화번호가 있고, 각각 13,625개, 273,920개의 관측치 존재

– 데이터셋에서 사업자등록번호 1개 당 웹사이트 주소가 1개인 레코드는 7,719개, 2개인 레코드는 2,614개, 3개인 레코드는 984개 등으로 총 13,453개의 레코드가 사업자등록번호와 연계되어 나타남

– 데이터셋에서 전화번호 1개 당 웹사이트 주소가 1개인 레코드는 179,582개, 2개인 레코드는 35,002개, 3개인 레코드는 10,131개 등으로 나타남

– 이는 1개 사업자등록번호 혹은 1개 전화번호에 웹사이트가 여럿 존재하는 케이스가 있음을 의미

○ (매칭) 1차 분석결과, 사업자등록번호를 이용해 데이터프로바이더에 기업등록부를 m:m 매칭했고, 기업등록부 11,434개 레코드가 매칭

- 이중 웹사이트 정보와 사업자등록번호가 1:1로 매칭된 것이 8,232개, m:1로 매칭된 것이 3,093개로 나타남

○ (매칭 전략) 기업등록부의 사업체에 웹사이트 분류를 적용하기 위해 5단계의 절차를 통해 매칭 분석

- (1단계) 데이터프로바이더 데이터셋과 기업등록부 데이터셋을 매칭. 이때, 1차적으로 사업자등록번호로 매칭하고, 2차적으로 전화번호로 매칭. 우선순위는 사업자등록번호, 전화번호 순서임
- (2단계) 1단계의 매칭 데이터를 기업등록부 id당 m개의 웹사이트 정보를 보유하도록 데이터 형태 전환. 이때 웹사이트 정보란 웹사이트 주소와 웹사이트 분류(카테고리 A~E)를 의미

※ 매칭 전략에 따르면, 기업등록부 id에 반복적인 매칭과 분석이 요구되며, 원활한 연구를 위해 기업등록부에 대한 온전한 접근이 필수적임

- (3단계) 2단계의 매칭 데이터를 기업등록부 id를 키변수로 기업등록부에 1:1 매칭
- (4단계) 기업등록부 id당 부여된 m개의 웹사이트 정보를 1개로 통합하는 기준을 마련하고, 이를 이용해 1개 기업등록부 id에 1개의 웹사이트 분류(카테고리 A~E)를 적용
- (5단계) 산업, 종사자규모, 웹사이트 분류 등을 기준으로 사업체수, 매출액, 종사자수 등 기초통계를 산출하고 필요에 따라 법인등록번호로 통합 등을 검토

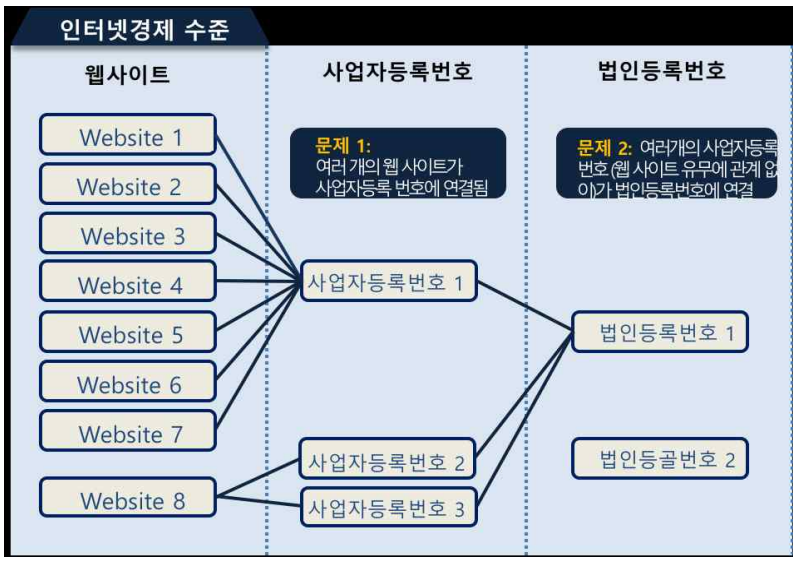
○ (분석단위) 웹사이트 분류를 위해서는 웹사이트 주소와 사업자등록번호와 법인등록번호를 분석단위로 삼아야 하지만, 데이터 매칭 및 주요통계 추출 과정에서는 기업등록부 데이터 특성상 통계청의 고유번호(기업등록부 id)를 분석단위로 이용

- 기업등록부에서 개인사업자가 여러 사업체를 운영하는 경우, 개별 사업체와 합계값이 레코드로 제공
- (i) 개별 사업체를 이용하는 경우는 아래 그림에서 웹사이트 - 사업자등록번

호까지만 이용하는 것이고,

- (ii) 개별 사업체들의 합계값을 이용하는 경우 아래 그림에서 법인등록번호까지 모두 이용하는 것임

〔그림 3-5〕 한국 데이터를 이용한 웹사이트와 사업자등록번호, 법인등록번호 간 관계



3. 데이터 프로바이더 데이터 특성[5)]

○ (사업체명) 데이터프로바이더 데이터에서 사업체 명이 있는 웹사이트는 14만 6천 개로 전체의 16%를 차지

○ (사업자등록번호) 사업자 등록번호가 기입된 웹사이트는 1만 3천 개로 전체의 1%에 불과함

○ (1차 매칭 주요결과) 데이터프로바이더를 기업등록부에 매칭한 1차 분석결과를 통해 아래와 같은 빈도표를 추출

5) 〔부록 1〕 데이터 프로바이더 데이터 변수 설명 참고

〈표 3－5〉 세부 유형별 매칭 주요 결과

구분	레코드 수	온라인 스토어	카트 시스템 보유	지불수단 정보	지불업체 정보
전체	925,961	2,017	10,226	3,801	4,587
매칭된 레코드	11,432*	55	226	85	87
매칭되지 않은 레코드	914,527	1,962	10,000	3,716	4,500

* 2개 레코드는 데이터 정제과정에서 삭제

○ (품질의 한계) 한국의 인터넷 크롤링 데이터 품질의 한계가 나타남
- 전반적으로 데이터 정제 상태가 미흡하며, 사업자등록번호만을 이용해 매칭하면 매칭 결과가 미흡(1%)한 문제 발생
- 매칭 범위 확대를 위해 보조 정보(전화번호, 사업체명 등) 활용 필요

○ (전화번호) 전화번호가 기입된 웹사이트는 27만 개로 전체의 30% 수준

4. 우리나라 데이터 적용을 위한 수정된 매칭 전략

○ (수정된 매칭전략) 데이터 프로바이더의 정보를 기업등록부에 매칭하기 위해 보완적 정보를 활용한 수정된 매칭 전략을 활용하기로 함

○ (1단계) 사업자등록번호를 이용해 데이터프로바이더 자료와 기업등록부 자료를 매칭
- 기업등록부 id당 m개의 웹사이트 정보 보유하도록 데이터를 정비
- id당 부여된 웹사이트 정보를 특정 기준에 맞춰 정리한 후, 분류(카테고리 A~E) 적용
- 웹사이트 정보가 복수로 존재시 B→C→D→E 순으로 분류를 적용함

○ (2단계) 외부자료(공정위, 온라인쇼핑몰, KED) 웹사이트 정보를 사용하여 데이터 프로바이더 사업자등록번호를 보완 후, 기업등록부와 추가 매칭
- 공정위와 온라인쇼핑몰 정보가 보완된 경우 카테고리 C 단계 이상으로 분류하고, KED 자료가 보완된 경우 카테고리 B 단계 이상으로 분류

○ (3단계) 공정위, 온라인쇼핑몰, KED의 사업자 등록번호를 이용해 기업등록부

에 추가 매칭

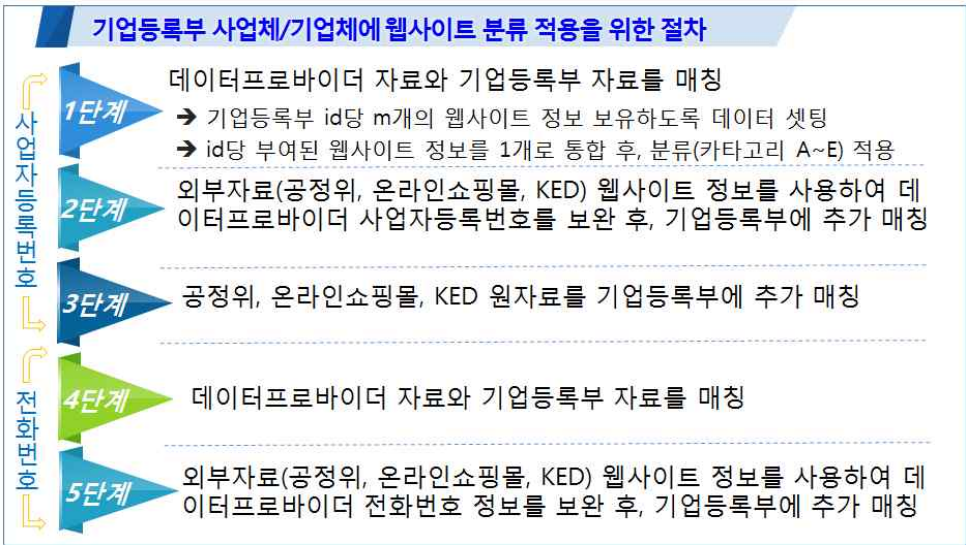
– 공정위와 온라인쇼핑몰의 경우 C 분류에 해당

– KED 자료의 경우 웹사이트가 존재하는 B 단계에 해당

○ (4단계) 전화번호를 이용해 데이터프로바이더 자료와 기업등록부 자료를 매칭

○ (5단계) 외부자료(공정위, 온라인쇼핑몰, KED) 웹사이트 정보를 사용하여 데이터 프로바이더 전화번호 정보를 보완 후, 기업등록부에 추가 매칭

〔그림 3－6〕 우리나라 데이터 적용 분석을 위한 매칭 전략



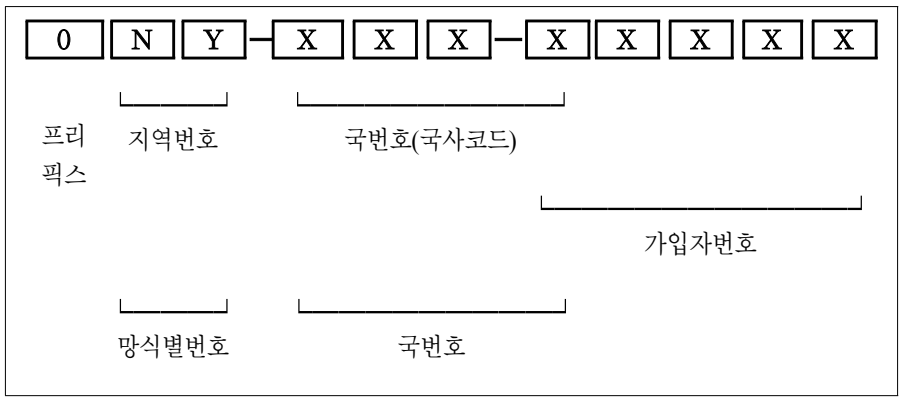
○ (전화번호 자료 전처리) 우리나라 번호체계는 프리픽스, 지역번호, 국번호, 가입자번호로 구성

– 데이터프로바이더, 공정위 “통신판매사업자 신고자료”, 통계청 “온라인쇼핑 동향조사 모집단 자료”의 전화번호를 아래 기준에 맞게 정비

– 자릿수가 맞지 않거나, 0000 등의 가입자 번호는 삭제 처리

○ (번호체계 소개) 통신망식별번호는 네트워크를 기준으로 제공서비스의 식별코드를, 지리적 번호는 시내전화(PSTN망)서비스가 제공되는 지역을 기준으로 코드를 구분

〔그림 3－7〕 우리나라 번호체계



※ 미래창조과학부는「전기통신사업법」제48조(전기통신번호자원 관리 계획)를 근거로, 전기통신역무의 효율적 제공과 이용자 편익 증대 및 공정경쟁 환경 조성 등을 위한 번호체계 및 관리계획 수립

○ 통신망식별번호는 국제전화망 ‘00Y’, 이동통신망은 ‘01Y’, 인터넷전화망은 ‘070’ 번호 대역을 할당하여 사업자에게 번호를 부여

－ 국제전화서비스는 앞서 설명한 번호체계와 다소 상이하게 Prefix ‘0’ 2자리와 국제전화망 식별코드가 조합된 행태이며, 대역별로 001~002, 005~006, 008번호는 기간통신사업자, ‘003, 007’은 별정통신사업자 용도로, 004, 009는 예비 대역으로 지정

－ 이동통신서비스는 Prefix ‘0’ 1자리와 국내착신번호 ‘1Y’가 조합된 행태이며, 010, 016, 017, 018, 019 대역별은 이동통신용, 012는 M2M용 013은 특정무선통신 대역으로 지정

－ 인터넷 전화번호는 Prefix ‘0’ 1자리와 국내착신번호 ‘70’이 조합된 번호를 부여하였고, 나머지 ‘7Y’계열은 예비전호로 지정

○ 그 중 이동통신서비스 식별번호로 지정된 01X계열의 011, 016, 017, 018, 019 등의 번호는 2002년 2월 정보통신부의 ‘010 번호통합정책’ 시행에 따라 번호회수가 결정되었으며, 현재까지 SKT 및 LGU+의 2G서비스에 부여된 011, 019번호는 회수 중에 있음

○ 다음으로, 지역번호는 시내전화 서비스(PSTN)에 할당하였고, 특별시, 광역시, 도 단위에 따라 ‘02Y, 03Y, 04Y, 05Y, 06Y’ 번호대역으로 구분

– 특별시 및 광역시 중 서울은 ‘02’, 인천광역시 ‘032’, 세종특별자치시 ‘044’, 부산광역시 ‘051’, 울산광역시 ‘052’, 대구광역시 ‘053’, 광주광역시 ‘062’로 부여

– 도 단위를 기준으로 경기도 ‘032’, 강원도 ‘033’, 충청남도 ‘041’, 충청북도 ‘043’, 경상북도 ‘054’, 경상남도 ‘055’, 전라남도 ‘061’, 전라북도 ‘063’, 제주도 ‘064’ 대역을 지정

○ 그밖에 착신과금 서비스 및 시외전화용 식별변호는 08Y의 번호대역을 사용하고 있으며, 09Y 번호대역은 예비자원으로 활용

〈표 3－6〉 통신망식별번호 및 지역번호 할당 현황

<table>
<tr><td>구분</td><td>0</td><td>1</td><td>2</td><td>3</td><td>4</td><td>5</td><td>6</td><td>7</td><td>8</td><td>9</td></tr>
<tr><td rowspan="2">00Y</td><td rowspan="2">예비</td><td colspan="8">국제전화용(통신망) 식별번호</td><td rowspan="2">예비
번호</td></tr>
<tr><td colspan="2">기간통신</td><td colspan="2">별정통신</td><td>예비</td><td colspan="2">기간통신</td><td>별정
통신</td><td>기간
통신</td></tr>
<tr><td rowspan="2">01Y</td><td colspan="10">이동통신용(통신망) 식별번호</td></tr>
<tr><td>IMT
(3G)</td><td>셀룰러
(2G)</td><td>M2M</td><td>특정무선통신
(항만, TRS 등)</td><td>무선
데이터
통신</td><td>무선
호출
(지역)</td><td>PCS(2
G)</td><td>셀룰
러
(2G)</td><td colspan="2">PCS
(2G)</td></tr>
<tr><td>02Y</td><td colspan="2">예비 번호</td><td colspan="8">서울</td></tr>
<tr><td>03Y</td><td>통합
메세징</td><td>경기</td><td>인천</td><td colspan="2">강원</td><td colspan="6">예비번호</td></tr>
<tr><td>04Y</td><td>예비
번호</td><td>충남</td><td>대전</td><td colspan="2">충북</td><td>세종</td><td colspan="5">예비번호</td></tr>
<tr><td>05Y</td><td>개인
번호</td><td>부산</td><td>울산</td><td colspan="2">대구</td><td>경북</td><td>경남</td><td colspan="4">예비번호</td></tr>
<tr><td>06Y</td><td>전화
정보</td><td>전남</td><td>광주</td><td colspan="2">전북</td><td>제주</td><td colspan="5">예비번호</td></tr>
<tr><td>07Y</td><td>인터넷
전화</td><td colspan="9">예비번호</td></tr>
<tr><td>08Y</td><td>착신
과금</td><td colspan="5">시외전화 기간통신</td><td>시외
별정</td><td>시외
기간</td><td colspan="3">예비번호</td></tr>
<tr><td>09Y</td><td colspan="10">예비 대역</td></tr>
</table>

자료: 「전기통신번호관리세칙」 제8조의 내용 재구성.

○ 한편, 주요 서비스별 번호체계는 아래 표와 같음

〈표 3－7〉 주요 서비스별 번호체계

서비스 구분	번호체계
이동전화(010)	010-AYYY-YYYY
사물지능통신(012)	012-AYYY-YYYY
무선호출(015)	015-XYY-YYYY 및 015-AYYY-YYYY
통합메시징서비스(030)	030-YYYYY-YYYY
개인번호서비스(050)	050-AYYY-YYYY, 050-AYYYY-YYYY
전화정보서비스(060)	060-AYY-YYYY
인터넷전화(070)	070-AYYY-YYYY
착신과금서비스(080)	080-YYY-YYYY
주파수공용통신서비스(0130)	0130-YYY-YYYY
선박무선통신(0131)	0131-YY-YYYY
무선데이터통신(0133)	0133-AYY-YYYY
위성휴대통신(0100)	0100-AY-YYYY

5. 카테고리 D와 E의 키워드 분석

○ (정의) 인터넷 분류 카테고리 D와 E의 업체는 다음과 같음

－ (카테고리 D) 플랫폼 사업자로 온라인서비스를 통해 매출을 발생시키지만, 소위 온라인스토어를 통해 매출을 발생시키지는 않는 업체

※ 데이트 사이트, 가격비교 사이트(booking.com), 뉴스 사이트 등

－ (카테고리 E) ICT 관련 인터넷 서비스를 통해 매출을 발생시키는 업체

○ (분류 방법) 카트 등의 유무로 일괄적인 분류가 불가능하므로, 웹사이트 크롤링 정보를 이용해 텍스트 분석을 수행

－ 키워드 선별 → 키워드 정제 → 조정

○ 네덜란드 인터넷 경제 분류의 카테고리 D는 온라인 서비스, E는 인터넷 관련 ICT로 구분됨

－ 인터넷 분류는 다시 세부 분류와 토픽으로 구분되어 있으며 토픽에 포함하는 주요 키워드를 추출해 아래 표와 같이 정리함

〈표 3－8〉 네덜란드 인터넷 경제 분류의 카테고리 D와 E

인터넷 분류	세분류	토픽	키워드: 기본	키워드: 결합
Category D: 온라인 서비스 (Online services)	레저 (Leisure)	숙박	호텔, 리조트, 호스텔, 침실 및 조식, 예약	호텔
		항공	비행기 표, 항공 상점, 티켓	항공편, 항공, 항공사
		여행	여행, 여행 전문가, 휴일, 휴가	온라인, 웹 사이트, 사이트, 비교
		음식	가정 배달, 음식, 먹다, 조리법, 식당, 요리	먹다. 온라인, 비교
	뉴스 및 엔터테인먼트 (News and entertainment)	뉴스	뉴스, 날씨, 미디어, 잡지	온라인
		블로그	블로그, 블로거	온라인
		빌로그	빌로그, 빌로거	온라인
		게임	게임, 놀이	온라인, 웹 사이트, 사이트, 무료
		비디오/음악	텔레비전, 영화, 시리즈, 클립, 비디오, 음악	온라인, 스트리밍, 스트림, 보기
		서적	서적	온라인
		e－러닝	학습, 코스, 배우다	온라인
		도박	도박, 카지노, 슬롯 머신	온라인
		성인	에스코트, 포르노, 섹스, 누드, 변태, 게이, 에로틱 한	온라인
	비즈니스 (Business)	광고	광고, 광고 대행사	온라인, 최적화, 애드워즈
		금융	금융, 투자, 온라인 투자, 투자자,	온라인, 비트 코인, 디지털
		컨설팅	조언, 컨설팅	온라인
		직업	공석 사이트, 구인 구직 사이트, 고용 기관, 임시 기관, 대리점 업무, 임시 근로자, 휴일 근로자, 휴가, 휴일 직업, 파트 타임 직업, 아르바이트, 직업, 채용, 신병 모집, 경력	온라인, 검색, 찾기
	소매 (Retail)	부동산	집, 주택, 소유자가 점령 한 부동산, 중개인, 임대 주택, 모기지	온라인, 검색, 찾기
		가격비교	에너지, 보험, 구독, 가격, 공급자, 저렴한, 저장, 자동차 보험, 가격 비교, 모바일	비교, 비교하다
		티켓	발권, 티켓	온라인
		경매	경매, 온라인 경매, 경매인, 인터넷 경매	온라인
		온라인거래	경매, 온라인 경매, 마켓 플레이스, 중고품	온라인

인터넷 분류	세분류	토픽	키워드: 기본	키워드: 결합
Category D: 온라인 서비스 (Online services)	일반 서비스(General services)	데이트	데이트, 섹스 데이트	온라인
		시각화	시각화, 애니메이션, 디자인	온라인, 3d
		운송	자동차, 리프트, 타기, 9292, 여행 플래너, 경로 플래너, 주차	온라인, 앱
		온라인 결제	서비스, 지불, 지불 솔루션	온라인
Category E: 인터넷 관련 ICT (Internet related ICT)	호스팅 및 클라우드(Hosting and cloud)	웹호스팅	호스팅, 도메인 이름, 웹 호스팅, 서버, 서비스, 클라우드, 개발,	웹 사이트, 스토리지, 클라우드, 온라인
		클라우드서비스, 데이터센터	클라우드 서비스, 데이터 센터, 클라우드 컴퓨팅, 대용량 저장 장치, 데이터 저장 장치, 데이터베이스, 데이터웨어 하우스, 데이터 처리, 발견, 전달, 사용, 회전율, 비즈니스, 공급 업체, 공급자	웹 사이트, 스토리지, 클라우드, 온라인
		웹 디자인 개발	웹 디자인, 인터넷 에이전시, 웹 사이트 구축, 웹 포털, 개발자, 웹 사이트, 디자인, 소프트웨어, 개발하다, 모바일, 개발자, 개발, 그래픽, 원래, 본, 운영, 이미지, 사운드, 마케팅, 양방향, 개인 정보 보호, 상호 작용하는, 움직이는, 이야기, 애니메이션, 멀티미디어, 빌더, 빌드, 생산자, 상인, 디자인	소프트웨어, 패키지, 가상, 온라인, 웹 사이트, 웹
		앱디자인 개발	모바일, 아이폰, 안드로이드, 태블릿, 과제, 출판	앱, 디지털, 애플리케이션
	소프트웨어 (Software)	소프트웨어 제품 및 서비스	소프트웨어, 개발, 제작, 게시, 조언, 데이터베이스	
	마케팅 및 컨설팅(Marketing and consultancy)	인터넷 마케팅	전자 상거래, 이마케팅, 애드워즈, 인터넷 마케팅, 검색 엔진, 온라인, 검색 엔진, b2b, b2c, 판매, 구매, 마케팅, 최적화, 소셜 미디어, 사회적, 분석, 광고주, 정보, 사생활	온라인, 웹 사이트, 전자, 인터넷, 네트워크, 데이터, 검색, 사업
		인터넷 컨설팅	인터넷, 온라인, 개발, 컨설팅, 연구, 조언, 분석, 기술 회사, 기술, 커뮤니케이션, 회계 공급자, 위키, 커뮤니티, 행 아웃, 스카이프, 메시징, 커뮤니케이션, 유튜브, 트윗, 트위터, 스냅 챠트, 페이스 북	인터넷, 컨설팅, 시간, 온라인, 디지털, 인터넷, 웹 서비스, 네트워크, IT, ICT,
	인프라 및 보안(Infrastructure and security)	방화벽	방화벽, 사이버, VPN, 스파이웨어, 안티 스팸, 네트워크 보안, 해킹, 해커, 인터넷 위협, 보안, 피싱, 추적, 파밍, 위험, 사건, 바이러스, 관리, 솔루션, 보호, 전문가, 설치, 침입, 절단, 암호, 신용 카드 정보, 보호, 봉쇄	응용 프로그램, 고급, 사이버, 웹, 인터넷, 웹 서비스, 네트워크, 데이터, 소유권, 공개, 안전

인터넷 분류	세분류	토픽	키워드: 기본	키워드: 결합
Category E: 인터넷 관련 ICT (Internet related ICT)	인프라 및 보안(Infrastructure and security)	데이터 마이닝 및 빅데이터	로봇, 자동화, 크롤러, 데이터 수집, 데이터 마이닝, 텍스트, 웹 디자인, 데이터, 공급 업체, 전문가, 적용, 구글, 고려, 발견, 패턴, 지능, 개발, 프로그래밍, 플랫폼, 인식, 기계, 수집, 지역, 회사 데이터, 수집, 과학	데이터, 인공, 학습, 통신, 소프트웨어, 패키지, 가상, 온라인, 웹 사이트, 앱, 애플리케이션, 디지털, 전자, 응용 프로그램, 고급, 스토리지, 클라우드, 인터넷, 컨설팅, 웹 서비스, 네트워크, 데이터, 보안, IT, ICT

○ 네덜란드 인터넷 경제 분류의 D와 E에 해당하는 국내 업체의 키워드 분석을 위해 국내 주요기업과 네덜란드 데이터간의 매칭 작업이 필요

○ 국내 주요 기업은 NICE평가정보원 kisvalue에서 제공하는 코스피, 코스탁 등 상장기업과 외감 기업을 포함한 약 3만개 기업 데이터 중 홈페이지가 존재하는 약 1만 개의 기업 데이터를 대상으로 함

- 국내 인터넷 산업 분류를 위해 표준산업분류(KSIC9)별 kisvalue의 데이터를 활용해 네덜란드 인터넷 경제 분류의 D와 E에 해당하는 산업 분류내 기업을 추출함
- 위 과정을 거쳐서 kisvalue 데이터베이스에서 인터넷 경제 분류의 카테고리 D와 E에 해당하는 산업 분류내 기업 1,336개 기업을 추출

○ (텍스트 분석: 키워드 추출) kisvalue 기업 데이터에서 추출한 카테고리 D & E에 해당하는 1,336개 주요 업체의 생산품목을 단어별로 파싱하여 텍스트 분석을 수행

- 1,336개 기업의 주요 생산물의 텍스트 분석을 통해 이중 D와 E에 해당하는 주요 키워드를 각각 47개, 38개를 선별

〈표 3－9〉 D&E에 속하는 키워드 종류

카테고리 D		카테고리 E	
광고	방송프로그램	웹호스팅	데이터베이스보안
광고대행	게임소프트웨어	호스팅솔루션	DRM솔루션
전자상거래	게임	헬스케어플랫폼	DB보안솔루션
온라인정보제공	모바일게임	푸시전송플랫폼	소프트웨어
온라인정보	온라인게임	클라우드잇서비스	자료처리
정보제공	전자책	클라우드	시스템통합자문
정보서비스	방송컨텐츠	콘텐츠관리시스템	시스템통합구축
전자지급결제	텔레비젼방송	서버호스팅	시스템통합관리
전자화폐	케이블방송	데이터센터	시스템컨설팅
관광호텔	이러닝	공공클라우드	시스템정보구축
호텔	웹툰	보안솔루션	시스템매니지먼트서비스
콘도미니엄	웹소설	정보보안솔루션	시맨틱웹솔루션
여행사	온라인게임소프트웨어	보안	스토리지앤드솔루션 프로그램
리조트	인터넷교육컨텐츠	보안컨설팅	
휴양콘도	모바일컨텐츠	보안프로그램	
여행	모바일게임퍼블리싱	보안장비	
콘도	모바일게임소프트웨어	보안시스템	
유스호스텔	모바일게임개발	해킹탐지	
관광숙박	전자출판	통합인증	
호텔예약	통신판매	통합보안패키지	
종합휴양콘도	인터넷쇼핑몰	통합보안솔루션	
펜션	쇼핑구매대행	침입탐지시스템	
여객운송		정보유출방지솔루션	
온천호텔		전자상거래보안솔루션	
출판		모의해킹	
총 47개		총 38개	

○ (텍스트 분석: 사업체 추출) 데이터 프로바이더의 description 정보에 주요 키워드가 존재하고, 통계청 BR의 산업분류상 국내 인터넷 경제 분류의 D와 E에 대한 표준산업분류에 해당하는 사업체를 인터넷 경제의 카테고리 D와 E에 포함
－ 텍스트분석 결과만을 사용하는 것은 분류의 정확성이 높지 않다고 판단하여, 기업등록부의 표준산업분류 정보를 동시에 적용하여 범위를 한정했음

– 즉, 사업체의 웹사이트 키워드가 D의 핵심 키워드를 포함하며, 사업체의 산업이 D로 정의한 산업에 포함되는 경우는 해당 사업체를 D로 분류

– E도 동일한 방법으로 분류함

○ D와 E에 해당하는 표준산업분류는 다음 방법으로 선정

– 정보서비스(63000), 시스템·응용 소프트웨어 개발 및 공급업(58220) 같이 다양한 토픽이 포함된 산업은 세세분류에 해당하는 산업까지 따로 분류하고, 분류가 되지 않은 기업들은 kisvalue에서 제공하는 주요 상품명에 대한 키워드 분석을 통해 세부 분류로 분류함

※ 예를 들어 정보서비스 내 포털 및 기타 인터넷 정보매개 서비스업(63120) 중 주요 상품 설명에 인터넷 광고를 포함하는 포털의 경우 세부 분류인 비즈니스의 광고 토픽으로 분류

※ 카테고리 E의 인터넷 관련 ICT는 시스템·응용 소프트웨어 개발 및 공급업(58220) 산업이 대다수를 차지하고 있는데 세세산업분류로 분류하기 어려워 주요 상품명에 클라우드, 웹, 마케팅, 보안 등의 키워드를 통해 세부 분류 및 토픽으로 분류함

〈표 3－10〉 국내 인터넷 경제 분류의 D와 E에 대한 표준산업분류

인터넷 분류	세부 분류	토픽	표준산업분류(KSIC9)	
Category D: 온라인 서비스 (Online services)	레저	숙박	55000	숙박업
			63120	포털 및 기타 인터넷 정보매개 서비스업
			63991	데이터베이스 및 온라인정보 제공업
		항공	51000	항공 운송업
			63991	데이터베이스 및 온라인정보 제공업
		여행	75200	여행사 및 기타 여행보조 서비스업
		음식	56100	음식점업
	뉴스 및 엔터테인먼트	뉴스	63120	포털 및 기타 인터넷 정보매개 서비스업
			63910	뉴스 제공업
			63991	데이터베이스 및 온라인정보 제공업
		게임	58210	게임 소프트웨어 개발 및 공급업
			63991	데이터베이스 및 온라인정보 제공업
			63990	그외 기타 정보 서비스업

인터넷 분류	세부 분류	토픽	표준산업분류(KSIC9)	
Category D: 온라인 서비스 (Online services)	뉴스 및 엔터테인먼트	비디오	59000	영상・오디오 기록물 제작 및 배급업
			60000	방송업
			63991	데이터베이스 및 온라인정보 제공업
		음악	63991	데이터베이스 및 온라인정보 제공업
		서적 및 e-러닝	58000	출판업
			63120	포털 및 기타 인터넷 정보매개 서비스업
			63114	자료 처리업
			63991	데이터베이스 및 온라인정보 제공업
	비즈니스	광고	71300	광고업
			63120	포털 및 기타 인터넷 정보매개 서비스업
			63991	데이터베이스 및 온라인정보 제공업
		금융	64000	금융업
		컨설팅	63120	포털 및 기타 인터넷 정보매개 서비스업
		직업	63120	포털 및 기타 인터넷 정보매개 서비스업
			63991	데이터베이스 및 온라인정보 제공업
		기타	63111	자료 처리업
			63120	포털 및 기타 인터넷 정보매개 서비스업
			63990	그외 기타 정보 서비스업
			63991	데이터베이스 및 온라인정보 제공업
	일반서비스	온라인 결제	63991	데이터베이스 및 온라인정보 제공업
			58220	시스템・응용 소프트웨어 개발 및 공급업
		운송	63991	데이터베이스 및 온라인정보 제공업
Category E: 인터넷 관련 ICT (Internet related ICT)	호스팅 및 클라우드	웹호스팅	63112	호스팅 및 관련 서비스업
		클라우드 서비스, 데이터센터	58220	시스템・응용 소프트웨어 개발 및 공급업
			62000	컴퓨터 프로그래밍, 시스템 통합 및 관리업
		웹사이트 디자인 개발	58220	시스템・응용 소프트웨어 개발 및 공급업
		앱 디자인 개발	58220	시스템・응용 소프트웨어 개발 및 공급업
	소프트웨어	소프트웨어 제품 및 서비스	58220	시스템・응용 소프트웨어 개발 및 공급업
	마케팅 및컨설팅	인터넷 마케팅	58220	시스템・응용 소프트웨어 개발 및 공급업
	인프라 및 보안		58220	시스템・응용 소프트웨어 개발 및 공급업

○ (추가 점검) 데이터 프로바이더의 description 정보에 주요 키워드가 존재하고, “online”, “온라인” 이라는 키워드가 교차검색되고, 데이터프로바이더 카테고리 항목이 특정 항목*에 해당하는 경우를 카테고리 D와 E에 해당한다고 파악

※ Adult, Entertainment, Finance, IT-services and Telecom, Leisure, Multimedia, Science and Education, Travel and Transport

– 텍스트 교차분석결과 카테고리 D 및 E로 분류된 웹사이트는 363개가 추출되었고, 웹사이트에 하나씩 접속하여 D와 E의 해당여부를 분석한 결과 총 5건만 D와 E에 해당하지 않는 것으로 나타남

- 사업체 아님 6건: 1.7%
- 접속불가 16건: 4.4%
- 카테고리 B로 분류변경 3건: 0.8%
- 카테고리 C로 분류변경 2건: 0.5%

6. 인터넷 사용에 따른 업체 분류

○ 인터넷 경제 측정을 위한 분류 기준을 정리하면 다음과 같음

○ 카테고리 A: 웹사이트 미보유. 인터넷으로 인한 소득 없음

– 데이터프로바이더 및 외부자료(공정위 “통신판매사업자 신고자료”, 통계청 “온라인쇼핑동향조사 모집단 자료”, “KED 웹사이트자료)에 연계되지 않은 사업체

○ 카테고리 B: 웹사이트 보유. 인터넷을 통해 간접적으로 소득 발생

– 데이터프로바이더에 연계되지만 카트가 없는 사업체

– KED 데이터와 연계된 사업체

○ 카테고리 C: 웹사이트 보유. 인터넷을 통해 직접 소득 발생 (온라인상점)

– 데이터프로바이더에 연계되고 카트 보유 사업체

– 공정위 “통신판매사업자 신고자료”, 통계청 “온라인쇼핑동향조사 모집단 자료”와 연계된 사업체

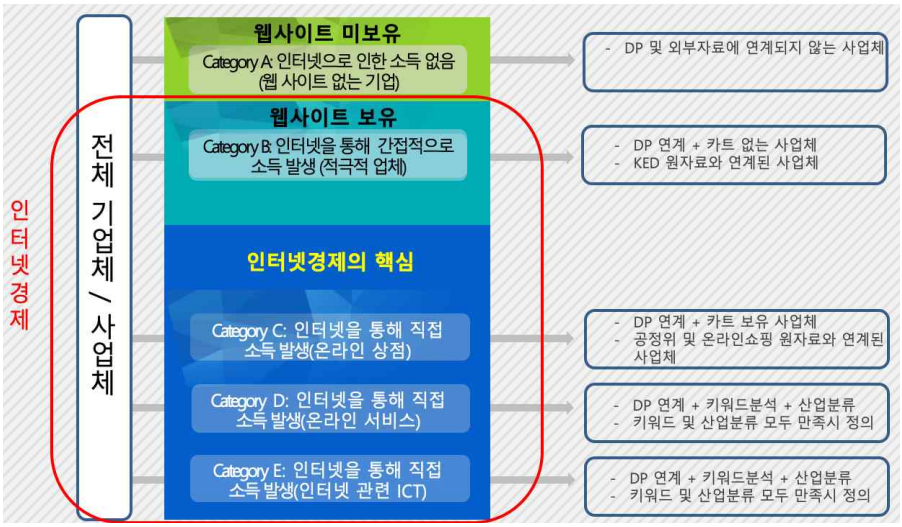
○ 카테고리 D: 웹사이트 보유. 인터넷을 통해 직접 소득 발생(온라인서비스)

– 카테고리 D 키워드 및 기업등록부의 산업분류를 모두 만족하는 사업체

○ 카테고리 E: 웹사이트 보유. 인터넷을 통해 직접 소득 발생(인터넷 관련 ICT 서비스)

– 카테고리 E 키워드 및 기업등록부의 산업분류를 모두 만족하는 사업체

〔그림 3－8〕 인터넷 사용에 따른 업체 분류



제 3 절 우리나라 인터넷 경제 측정 결과

1. 데이터 단계별 매칭 결과

○ (기업체 및 사업체 수) 데이터프로바이더(DP) 및 외부자료(공정위 “통신판매사업자 신고자료”, 통계청 “온라인쇼핑동향조사 모집단 자료”, “KED 웹사이트자료)를 사용하여 기업등록부(BR)에 매칭한 결과는 다음과 같음

○ 각 단계의 결과는 이전 단계의 결과를 포함한 누적 값을 나타냄

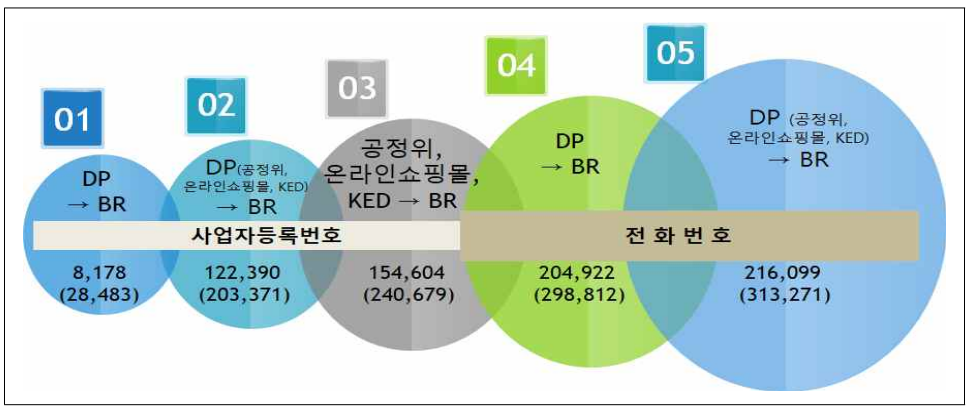
– (1단계) 데이터프로바이더의 사업자등록번호를 기업등록부에 연계한 결과, 8,178개의 기업체(28,483개의 사업체)가 매칭

– (2단계) 외부자료를 통해 데이터프로바이더의 사업자등록번호를 보완 후 기

업등록부에 연계한 결과, 122,390개의 기업체(203,371개의 사업체)가 추가로 매칭

- (3단계) 외부자료의 사업자등록번호를 직접 사용하여 기업등록부에 연계한 결과, 154,604개의 기업체(240,679개의 사업체)가 추가로 매칭
- (4단계) 데이터프로바이더의 전화번호를 기업등록부에 연계한 결과, 204,922개의 기업체(298,812개의 사업체)가 추가로 매칭
- (5단계) 외부자료를 통해 데이터프로바이더의 전화번호를 보완 후 기업등록부에 연계한 결과, 216,099개의 기업체(313,271개의 사업체)가 최종적으로 매칭됨

〔그림 3-9〕 데이터 매칭 단계별 기업체 및 사업체 수



○ (매출액 및 종사자) 연계된 데이터의 기업체 기준 매출액 및 종사자 수는 다음과 같음

- (1단계) 연계된 8,178개 기업의 총 매출액은 1,214조 원이며 종사자 수는 1,873,121명으로 나타남
- (2단계) 연계된 122,390개 기업의 총 매출액은 2,682조 원이며 종사자 수는 5,855,072명으로 나타남
- (3단계) 연계된 154,604개 기업의 총 매출액은 2,720조 원이며 종사자 수는 5,974,356명으로 나타남

– (4단계) 연계된 204,922개 기업의 총 매출액은 2,828조 원이며 종사자 수는 6,423,199명으로 나타남

– (5단계) 연계된 216,099개 기업의 총 매출액은 2,837조 원이며 종사자 수는 6,467,073명으로 나타남

– 각 단계의 결과값이 기업등록부 전체 값에서 차지하는 비중은 매출액 기준의 경우 20.6%(1단계) – 45.5%(2단계) – 46.2%(3단계) – 48.0%(4단계) – 48.2%(5단계)이며, 종사자 기준의 경우 9.2%(1단계) – 28.7%(2단계) – 29.2%(3단계) – 31.4%(4단계) – 31.7%(5단계)로 나타났음

〈표 3–11〉 데이터 매칭 단계별 매출액 및 종사자 수

구분	매출액 (조 원)	비중(%)	종사자수 (명)	비중(%)
기업등록부(BR) 전체	5,892	–	20,430,684	–
1단계(사업자등록번호 기준, DP → BR)	1,214	20.6	1,873,121	9.2
2단계(사업자등록번호 기준, DP + 외부자료 → BR)	2,682	45.5	5,855,072	28.7
3단계(사업자등록번호 기준, 외부자료 → BR)	2,720	46.2	5,974,356	29.2
4단계(전화번호 기준, DP → BR)	2,828	48.0	6,423,199	31.4
5단계(전화번호 기준, DP + 외부자료 → BR)*	2,837	48.2	6,467,073	31.7

2. 연계 데이터 출처별 매칭 결과

○ (기업체 및 사업체 수) 기업등록부와 연계된 데이터 출처별 매칭 결과는 다음과 같음

– (1그룹) 데이터프로바이더 원자료에 존재하는 사업자등록번호 및 외부자료를 통해 데이터프로바이더에 보완한 사업자등록번호를 기반으로 연계할 경우, 122,390개의 기업체(203,371개 사업체)가 매칭되는 것으로 나타남

※ 이 중 데이터프로바이더 원자료에서 8,178개 기업체(28,483개 사업체)가 연계되었으며, 외부자료를 통해 사업자등록번호를 보완한 경우에서 114,212개 기업체(174,888개 사업체)가 매칭되었음

– (2그룹) 1그룹에서 연결된 자료를 제외하고 외부자료의 사업자등록번호를 사

용하여 연계할 경우, 32,214개의 기업체(37,308개 사업체)가 매칭됨

- (3그룹) 1, 2그룹에서 연결된 자료를 제외 후 데이터프로바이더 원자료에 존재하는 전화번호 및 외부자료를 통해 데이터프로바이더에 보완한 전화번호를 기반으로 연계하면, 61,495개 기업체(72,592개 사업체)가 연결됨

※ 이 중 데이터프로바이더 원자료에서 50,318개 기업체(58,133개 사업체)가 연계되었으며, 외부자료를 통해 전화번호를 보완한 경우에서 11,177개 기업체(14,819개 사업체)가 매칭됨

〈표 3－12〉 연계 데이터 출처별 기업체 및 사업체 수

구분	기업체수(개)	비중(%)	사업체수(개)	비중(%)
기업등록부와 연계된 총 관측치	216,099	100.0	313,271	100.0
(1그룹) 사업자등록번호 기준 BR-DP	122,390	56.6	203,371	64.9
BR - DP 원자료	8,178	3.7	28,483	9.1
BR - 외부자료 통해 보완한 DP자료	114,212	52.9	174,888	55.8
(2그룹) 사업자등록번호 기준 BR-외부자료	32,214	14.9	37,308	11.9
(3그룹) 전화번호 기준 BR-DP	61,495	28.5	72,592	23.2
BR - DP 원자료	50,318	23.3	58,133	18.5
BR - 외부자료 통해 보완한 DP자료	11,177	5.2	14,819	4.7

○ (매출액 및 종사자 수) 연계된 데이터 출처별로 기업체 기준 매출액 및 종사자 수는 다음과 같음

- (1그룹) 연계된 122,390개 기업의 총 매출액은 2,682조 원이며 종사자 수는 5,855,072명으로 나타났음
- (2그룹) 연계된 32,214개 기업의 총 매출액은 37조 원이며 종사자 수는 119,284명으로 나타났음
- (3그룹) 연계된 61,495개 기업의 총 매출액은 118조 원이며 종사자 수는 492,717명으로 나타났음

〈표 3－13〉 연계 데이터 출처별 매출액 및 종사자 수

구분	매출액(조 원)	비중(%)	종사자 수(명)	비중(%)
기업등록부와 연계된 총 관측치	2,837	100.0	6,467,073	100.0
(1그룹) 사업자등록번호 기준 BR-DP	2,682	94.5	5,855,072	90.5
BR - DP 원자료	1,214	42.8	1,873,121	29.0
BR - 외부자료 통해 보완한 DP자료	1,468	51.7	3,981,951	61.5
(2그룹) 사업자등록번호 기준 BR-외부자료	37	1.3	119,284	1.8
(3그룹) 전화번호 기준 BR-DP	118	4.2	492,717	7.7
BR - DP 원자료	109	3.8	448,843	7.0
BR - 외부자료 통해 보완한 DP자료	9	0.4	43,874	0.7

3. 인터넷 단계별 매칭 결과

○ (주요결과) 매칭을 통해 인터넷 경제 기업으로 정의된 기업체의 특성은 다음과 같음

－ (분포) 인터넷 경제에 속한 기업체는 총 216,099개이며, 그룹별 분포를 살펴보면 B그룹 46%(99,391개), C그룹 53%(114,619개), D그룹 0.9%(1,916개), E그룹 0.1%(173개)로 나타났음

－ (매출액) 인터넷 경제에 속한 기업체의 총 매출액은 2,837조 원이며, B그룹이 70.9%(2,011조 원), C그룹이 28.6%(811조 원), D그룹이 0.5%(13조 원), E그룹이 0.1%(2조 원)을 차지하는 것으로 나타났음

－ (종사자) 인터넷 경제에 속한 기업체의 총 종사자 수는 6,467,073명이며, B그룹이 70.8%(4,580,103명), C그룹이 28.3%(1,832,580명), D그룹이 0.7%(46,359명), E그룹이 0.1%(8,031명)을 차지하는 것으로 나타났음

〈표 3－14〉 인터넷 단계별 기업체 수, 매출액 및 종사자 수

구분	기업체수	분포(%)	매출액(조원)	분포(%)	종사자수(명)	분포(%)
인터넷 경제 기업	216,099	100.0	2,837	100.0	6,467,073	100.0
B그룹	99,391	46.0	2,011	70.9	4,580,103	70.8
C그룹	114,619	53.0	811	28.6	1,832,580	28.3
D그룹	1,916	0.9	13	0.5	46,359	0.7
E그룹	173	0.1	2	0.1	8,031	0.1

4. 지역별, 산업별 분포 결과

○ (지역별 분포 결과) 기업등록부 전체 기업과 인터넷 경제 기업의 지역별 분표 비교 결과는 다음과 같음

– 기업등록부 전체 기업을 대상으로 살펴본 결과 서울(21.9%), 경기도(23.8%), 부산(6.6%), 경상남도(6.4%) 순으로 위치해 있는 것을 확인

– 인터넷 경제 기업의 경우 서울(36.8%), 경기도(26.3%), 부산(5.0%) 순으로 분포가 높은 것을 확인

– 분포에 따른 지역 순위는 유사하나, 인터넷 기업이 서울과 경기 등 수도권에 밀집된 정도가 상대적으로 높은 것을 볼 수 있음

〈표 3－15〉 전체기업 및 인터넷기업 지역별 분포 비교

구분	전체		인터넷 경제	
	기업체수	분포(%)	기업체수	분포(%)
서울특별시	1,576,832	21.9	79,400	36.8
부산광역시	471,857	6.6	10,859	5.0
대구광역시	328,629	4.6	7,746	3.6
인천광역시	367,989	5.1	10,156	4.7
광주광역시	193,461	2.7	3,609	1.7
대전광역시	195,433	2.7	4,981	2.3
울산광역시	141,316	2.0	2,244	1.0
세종특별자치시	30,938	0.4	423	0.2
경기도	1,708,766	23.8	56,605	26.3

구분	전체		인터넷 경제	
	기업체수	분포(%)	기업체수	분포(%)
강원도	220,329	3.1	4,480	2.1
충청북도	218,014	3.0	4,331	2.0
충청남도	284,047	3.9	5,687	2.6
전라북도	252,253	3.5	4,191	1.9
전라남도	262,275	3.6	3,760	1.7
경상북도	371,735	5.2	6,655	3.1
경상남도	462,923	6.4	8,141	3.8
제주특별자치도	105,576	1.5	2,311	1.1
전체	7,192,373*	100.0	215,579*	100.0

* 지역 정보가 없는 관측치는 제외한 수치

○ (산업별 분포 결과) 기업등록부 전체 기업과 인터넷 경제 기업의 산업별 분표 비교 결과는 다음과 같음

- 기업등록부 전체 기업의 경우 도매 및 소매업(23.3%), 부동산업 및 임대업(18.8%), 숙박 및 음식점업(12.0%) 순으로 기업체 수가 많은 것을 확인
- 인터넷 경제 기업의 경우 도매 및 소매업(34.3%), 제조업(21.4%), 출판, 영상, 방송통신 및 정보서비스업(8.1%) 순으로 분포가 높은 것을 확인
- 인터넷 경제 기업의 경우 도매 및 소매업의 비중이 높으며, 특히 제조업의 비중이 높은 것이 특징
- 지역별 분포보다는 산업별 분포에서 인터넷 경제 기업의 특징이 더 명확하게 드러나는 것으로 판단

〈표 3－16〉 전체기업 및 인터넷기업 산업별 분포 비교

구분	전체		인터넷	
	기업체수	분포	기업체수	분포
농업, 임업 및 어업	76,758	1.0	1,492	0.7
광업	2,660	0.0	71	0.0
제조업	558,983	7.8	45,150	21.4
전기, 가스, 증기 및 수도사업	19,756	0.3	261	0.1

구분	전체		인터넷	
	기업체수	분포	기업체수	분포
하수・폐기물 처리, 원료재생 및 환경복원업	9,816	0.1	361	0.2
건설업	440,168	6.1	7,740	3.7
도매 및 소매업	1,678,697	23.3	72,173	34.3
운수업	586,531	8.1	4,795	2.3
숙박 및 음식점업	863,765	12.0	10,912	5.2
출판, 영상, 방송통신 및 정보서비스업	109,318	1.5	16,993	8.1
금융 및 보험업	58,221	0.8	1,610	0.8
부동산업 및 임대업	1,351,536	18.8	12,221	5.8
전문, 과학 및 기술 서비스업	192,189	2.7	12,944	6.1
사업시설관리 및 사업지원 서비스업	112,746	1.6	5,978	2.8
공공행정, 국방 및 사회보장 행정	2,273	0.0	88	0.0
교육 서비스업	213,060	3.0	788	0.4
보건업 및 사회복지 서비스업	201,940	2.8	5,261	2.5
예술, 스포츠 및 여가관련 서비스업	143,508	2.0	2,724	1.3
협회 및 단체, 수리 및 기타 개인 서비스업	584,184	8.1	9,108	4.3
가구내고용활동및 달리 분류되지 않은 자가소비 생산활동	317	0.0	7	0.0
국제 및 외국기관	162	0.0	12	0.0
소계*	7,206,588	100.0	210,689	100.0

* 산업 정보가 없는 관측치는 제외한 수치

5. 인터넷 단계별 주요 변수 분포 결과

○ 기업등록부 전체 기업을 대상으로 인터넷 단계별 종사자 규모별, 인터넷 단계별 업력별, 인터넷 단계별 지역별, 인터넷 단계별 산업별로 구분하여 분포를 분석

– 인터넷 경제 기업을 나타내는 B~E그룹에 포함되지 않은 모든 기업은 A로 정의

○ (종사자규모 결과) 인터넷 단계별 및 종사자 규모별로 구분한 결과는 다음과 같음

– 종사자 규모 구간은 0~1명, 2~10명, 11~100명, 100명 초과로 정의

– 대부분의 인터넷 단계에서 0~1명 종사자 규모에 속한 기업의 숫자가 많음

을 알 수 있으며, 반대로 100명 초과 종사자를 보유한 기업의 비중이 낮음을 확인

- A그룹의 경우 0~1명 종사자를 보유한 소규모 기업체의 비중이 85.8%를 차지하며, C그룹(64.5%)과 D그룹(67.3%)에서도 소규모 기업체의 비중이 뚜렷함
- B그룹과 E그룹의 경우 상대적으로 종사자 규모에 따른 기업체 분포가 고른 것으로 판단

〈표 3－17〉 인터넷 단계별 종사자 규모별 기업 분포

구분	0~1명	2~10명	11~100명	100명 초과	소계
A	6,039,173 (85.8%)	797,855 (11.3%)	188,055 (2.7%)	12,297 (0.2%)	7,037,380
B	33,540 (33.8%)	28,848 (29.0%)	30,536 (30.7%)	6,467 (6.5%)	99,391
C	73,972 (64.5%)	27,108 (23.7%)	11,868 (10.3%)	1,671 (1.5%)	114,619
D	1,289 (67.3%)	391 (20.4%)	188 (9.8%)	48 (2.5%)	1,916
E	39 (22.5%)	54 (31.2%)	62 (35.9%)	18 (10.4%)	173
소계	6,148,013 (84.7%)	854,256 (11.8%)	230,709 (3.2%)	20,501 (0.3%)	7,253,479

○ (업력 결과) 인터넷 단계별 및 업력으로 구분한 결과는 다음과 같음

- 기업의 창설년도를 사용하여 업력을 계산하였으며 업력 구간은 0~1년, 2~5년, 6~10년, 10년 초과로 정의
- 업력의 경우 종사자 규모 대비 상대적으로 고른 분포를 보이고 있음을 확인
- B그룹과 E그룹의 경우, 10년 넘게 영업한 기업체의 수가 많은 것으로 나타남
- A그룹, C그룹, D그룹은 업력이 5년 이하인 기업체가 해당 그룹에 속한 기업의 절반 이상을 차지

〈표 3－18〉 인터넷 단계별 업력별 기업 분포

구분	0~1년	2~5년	6~10년	10년 초과	소계
A	1,873,094 (26.6%)	1,943,561 (27.6%)	1,291,612 (18.4%)	1,929,113 (27.4%)	7,037,380
B	12,244 (12.3%)	23,223 (23.4%)	20,815 (20.9%)	43,109 (43.4%)	99,391
C	27,741 (24.2%)	38,548 (33.6%)	24,892 (21.7%)	23,438 (20.5%)	114,619
D	506 (26.4%)	704 (36.7%)	382 (20.0%)	324 (16.9%)	1,916
E	22 (12.7%)	37 (21.4%)	43 (24.9%)	71 (41.0%)	173
소계	1,913,607 (26.4%)	2,006,073 (27.7%)	1,337,744 (18.4%)	1,996,055 (27.5%)	7,253,479

○ (지역 결과) 인터넷 단계별 및 지역으로 구분한 결과는 다음과 같음

－ 대부분의 지역에서 A그룹, 즉 비인터넷 경제 기업의 분포가 뚜렷함을 확인

－ 인터넷 기업 분포가 높은 서울과 경기도의 경우, B그룹과 C그룹의 비중이 높음을 확인

－ D그룹의 경우 서울, 경기도, 강원도, 경상남도, 충청남도 등에 주로 위치한 것으로 나타남

－ E그룹은 서울에 밀집되어 있음

〈표 3－19〉 인터넷 단계별 지역별 기업 분포

구분	A	B	C	D	E	소계
서울특별시	1,497,432	34,157	44,669	465	109	1,576,832
	(94.96%)	(2.17%)	(2.83%)	(0.03%)	(0.01%)	
부산광역시	460,998	5,491	5,310	54	4	471,857
	(97.70%)	(1.16%)	(1.13%)	(0.01%)	(0.00%)	
대구광역시	320,883	3,634	4,082	27	3	328,629
	(97.64%)	(1.11%)	(1.24%)	(0.01%)	(0.00%)	
인천광역시	357,833	4,777	5,280	92	7	367,989
	(97.24%)	(1.30%)	(1.43%)	(0.03%)	(0.0%)	
광주광역시	189,852	1,592	2,004	13		193,461
	(98.13%)	(0.82%)	(1.04%)	(0.01%)	(0.00%)	
대전광역시	190,452	2,502	2,467	8	4	195,433
	(97.45%)	(1.28%)	(1.26%)	(0.01%)	(0.00%)	

구분	A	B	C	D	E	소계
울산광역시	139,072	1,258	968	18		141,316
	(98.41%)	(0.89%)	(0.69%)	(0.01%)	(0.00%)	
세종특별자치시	30,515	276	146		1	30,938
	(98.63%)	(0.89%)	(0.48%)	(0.00%)	(0.00%)	
경기도	1,652,161	27,296	28,944	330	35	1,708,766
	(96.69%)	(1.60%)	(1.69%)	(0.02%)	(0.00%)	
강원도	215,849	1,645	2,604	229	2	220,329
	(97.97%)	(0.75%)	(1.18%)	(0.10%)	(0.00%)	
충청북도	213,683	2,230	2,067	30	4	218,014
	(98.01%)	(1.02%)	(0.96%)	(0.01%)	(0.00%)	
충청남도	278,360	2,767	2,787	133		284,047
	(98.00%)	(0.97%)	(0.98%)	(0.05%)	(0.00%)	
전라북도	248,062	1,785	2,348	58		252,253
	(98.34%)	(0.71%)	(0.93%)	(0.02%)	(0.00%)	
전라남도	258,515	1,609	2,094	56	1	262,275
	(98.57%)	(0.61%)	(0.80%)	(0.02%)	(0.00%)	
경상북도	365,080	3,155	3,389	111		371,735
	(98.21%)	(0.85%)	(0.91%)	(0.03%)	(0.00%)	
경상남도	454,782	4,326	3,632	180	3	462,923
	(98.24%)	(0.93%)	(0.79%)	(0.04%)	(0.00%)	
제주특별자치도	103,265	871	1,335	105		105,576
	(97.81%)	(0.82%)	(1.27%)	(0.10%)	(0.00%)	
소계*	6,976,794	99,371	114,126	1,909	173	7,192,373
	(97.00%)	(1.38%)	(1.59%)	(0.03%)	(0.00%)	

* 지역 정보가 없는 관측치는 제외한 수치

○ (산업 결과) 인터넷 단계별 및 산업으로 구분한 결과는 다음과 같음

- B, C그룹의 경우 제조업, 도매 및 소매업, 출판, 영상, 방송통신 및 정보서비스업에 주로 분포되어 있음
- D그룹 기업은 숙박 및 음식점업, 사업시설 관리 및 사업지원 서비스업에 많은 수가 분포되어 있음
- E그룹의 대부분은 출판, 영상, 방송통신 및 정보서비스업로 분류되고 있음을

확인

– 특정 산업 중심으로 인터넷 경제 기업이 분포되어 있으며, 서비스업이 인터넷 경제에서 차지하는 중요성이 높은 것으로 판단

〈표 3－20〉 인터넷 단계별 산업별 기업 분포

구분	A	B	C	D	E	소계
농업, 임업 및 어업	75,266	409	1,083			76,758
	(98.06%)	(0.53%)	(1.41%)	(0.00%)	(0.00%)	
광업	2,589	40	31			2,660
	(97.33%)	(1.50%)	(1.17%)	(0.00%)	(0.00%)	
제조업	513,833	29,584	15,557	7	2	558,983
	(91.92%)	(5.29%)	(2.79%)	(0.00%)	(0.00%)	
전기, 가스, 증기 및 수도사업	19,495	142	118	1		19,756
	(98.67%)	(0.72%)	(0.60%)	(0.01%)	(0.00%)	
하수・폐기물 처리, 원료재생 및 환경복원업	9,455	275	86			9,816
	(96.32%)	(2.80%)	(0.88%)	(0.00%)	(0.00%)	
건설업	432,428	4,899	2,834	7		440,168
	(98.24%)	(1.11%)	(0.65%)	(0.00%)	(0.00%)	
도매 및 소매업	1,606,524	20,809	51,338	16	10	1,678,697
	(95.70%)	(1.24%)	(3.06%)	(0.00%)	(0.00%)	
운수업	581,736	1,748	3,045	1	1	586,531
	(99.18%)	(0.30%)	(0.52%)	(0.00%)	(0.00%)	
숙박 및 음식점업	852,853	2,886	6,894	1,132		863,765
	(98.74%)	(0.33%)	(0.80%)	(0.13%)	(0.00%)	
출판, 영상, 방송통신 및 정보서비스업	92,325	8,312	8,273	252	156	109,318
	(84.46%)	(7.60%)	(7.57%)	(0.23%)	(0.14%)	
금융 및 보험업	56,611	936	669	5		58,221
	(97.23%)	(1.61%)	(1.15%)	(0.01%)	(0.00%)	
부동산업 및 임대업	1,339,315	5,034	7,171	16		1,351,536
	(99.10%)	(0.37%)	(0.53%)	(0.00%)	(0.00%)	
전문, 과학 및 기술 서비스업	179,245	7,258	5,587	96	3	192,189
	(93.26%)	(3.78%)	(2.91%)	(0.05%)	(0.00%)	
사업시설관리 및 사업지원 서비스업	106,768	2,218	3,383	377		112,746
	(94.70%)	(1.97%)	(3.00%)	(0.33%)	(0.00%)	

구분	A	B	C	D	E	소계
공공행정, 국방 및 사회보장 행정	2,185	59	29			2,273
	(96.13%)	(2.60%)	(1.27%)	(0.00%)	(0.00%)	
교육 서비스업	207,272	2,833	2,954		1	213,060
	(97.28%)	(1.33%)	(1.39%)	(0.00%)	(0.00%)	
보건업 및 사회복지 서비스업	196,679	4,747	514			201,940
	(97.39%)	(2.35%)	(0.26%)	(0.00%)	(0.00%)	
예술, 스포츠 및 여가관련 서비스업	140,784	1,232	1,488	4		143,508
	(98.10%)	(0.86%)	(1.04%)	(0.00%)	(0.00%)	
협회 및 단체, 수리 및 기타 개인 서비스업	575,076	5,721	3,385	2		584,184
	(98.44%)	(0.98%)	(0.58%)	(0.00%)	(0.00%)	
가구내고용활동및 달리 분류되지 않은 자가소비 생산활동	310	3	4			317
	(97.79%)	(0.95%)	(1.26%)	(0.00%)	(0.00%)	
국제 및 외국기관	150	12	0			162
	(92.59%)	(7.41%)	(0.00%)	(0.00%)	(0.00%)	
소계*	6,990,899	99,157	114,443	1,916	173	7,206,588
	(97.01%)	(1.37%)	(1.59%)	(0.03%)	(0.00%)	

* 산업 정보가 없는 관측치는 제외한 수치

제 4 절 네덜란드 방법론에 대한 한계점

1. 연구의 한계점

○ 네덜란드 측정방법론에 따른 인터넷 경제에 대한 규모 추정은 경제 내의 생산 구조를 파악하는데 기여할 것으로 보이나,

– 파일럿 연구의 특성상 다년도의 변화를 파악하는데 한계가 있어 단년도의 결과만으로는 의미 해석이 제한적이며

– 인터넷 경제 현상에 대한 이해와 정책적 함의 도출을 위한 국가(네덜란드와 한국)간 비교가 대상국의 크롤링 데이터 상황과 통계청 데이터와의 연계 정도 등이 상이함에 따라 제약이 존재

○ 네덜란드 측정 방법론은 한 국가의 인터넷 경제 규모를 측정하기 위한 파일럿 연구의 성격이 큼

– 이에 따라 언어의 차이, 연구의 기반이 되는 데이터의 활용가능성 등에 따라 인터넷 경제의 측정 규모가 상이할 여지가 큼
– 대부분의 기존 연구가 거시/산업 데이터인 국민계정(SNA) 데이터와 산업분류를 이용한 점을 고려할 때, 미시데이터를 연계하는 네덜란드 측정방법론은 연구 수행 방법론과 결과 해석에서 다양한 이슈를 야기할 것으로 보임
– 특히 네덜란드 측정방법론의 적용에서 가장 중요한 부분인 크롤링 데이터의 가용성 관점에서, 데이터프로바이더가 제공하는 데이터의 제약을 해소하기 위한 대안이 필요

○ 우리나라 현실에 적합한 인터넷 경제 규모 측정을 위해서는 글로벌 경쟁력을 통해 대규모의 수출이 발생하고 있는 ICT 기기 부문에 대한 추가적인 고려가 필요

2. 인터넷 경제/플랫폼 경제 규모에 대한 측정방법론 검토

○ 인터넷 경제/플랫폼 경제의 측정이 중요한 이유는 최근의 경제구조가 이전과는 다른 양상을 보이기 때문임
– 예를 들어, 인터넷을 통한 정보 활용 확산의 효과가 통계조사가 되는 생산측면으로 귀속되는 것 보다 조사되지 않는 소비측면으로 직접 파급
– 하지만, 인터넷 경제의 개념 정의와 이를 측정하기 위한 방법에 대한 공감대가 부족한 상황이며, 인터넷 경제와 플랫폼 경제에 대한 정의, 포함 관계도 명확하지 않음
– 예를 들어 한 기업체가 다양한 유형의 온라인, 오프라인을 통한 사업을 영위하고 있을 경우 이를 나누어 측정하는 것도 어려우며 하나의 분류로 할당하는 것도 어려운 상황임

※ 네덜란드 적용 사례에서도 은행은 각종 예적금 상품에 관한 정보를 온라인으로 제공하고 있으며(카테고리 B1), 카드를 온라인으로 신청하여 발급 받을 수도 있고(카테고리 B2), 인터넷에서 적금상품을 직접 가입할 수 있어(온라인 뱅킹, 카테고리 D) 다양한 속성을 지니고 있으나 인터넷 사업자 유형에서 B2로 구분하고 있음

○ 인터넷 경제의 측정을 위해서는 이를 측정할 수 있는 조작적 정의 및 범위 설정을 위한 정량적 측정이 요구됨

– (대안1) 크롤링 데이터 활용: 네덜란드 통계청에서 적용한 방법론을 이용하되, 국내의 웹크롤링 전문 업체를 통해 크롤링 데이터를 확보하여 통계청 데이터를 결합, 인터넷경제 규모를 추정

※ 대규모 웹 크롤링과 분석이 선행되어야 하기 때문에, 가용예산을 고려하여 조사범위를 한정하는 등 세부 수행 방안 마련이 필요(예를 들어 기업체 정보와 웹사이트가 연계된 데이터셋을 이용하여 크롤링 대상을 한정하는 방안 등)

– (대안2) 기존 통계자료 활용: 인터넷경제와 플랫폼 경제의 정의 및 포함관계 분석을 통해, 광고시장 규모를 통한 플랫폼경제 규모 추정. 또는 기존 정보화 통계조사 등을 통해 산업별 “네트워크에 연결된 컴퓨터 보유 사업체” 비율 등을 활용해 인터넷경제/플랫폼경제의 규모를 추정

○ 인터넷경제와 플랫폼경제의 다면적인 현황을 파악하기 위해서는 다음과 같은 다양한 관점의 통계조사 결과 활용이 필요

– 공급측면: 사업체의 생산액, 재고액, 종사자 수, 사업체 수

– 수요측면: 제품의 매출액, 수출액

– 인프라측면: 경제전반 혹은 기업 관점의 ICT, 인터넷, 데이터 등의 인프라에 대한 투자

– 이용자측면: 인터넷이용, 컴퓨터 보유, 스마트폰 보유, 정보격차 등

제 4 장 플랫폼 기업(비즈니스) 경제 규모에 대한 측정방법론 연구

제 1 절 연구의 필요성

○ 인터넷은 양적으로 끊임없이 확장, 성장해왔을 뿐 아니라 그 구조와 사용방식 등 질적인 측면에서도 크게 변화하고 진화하는 과정을 거듭해왔음

– 이러한 인터넷의 역동적인 특성을 고려할 때, 인터넷 경제의 개념을 시장 환경의 변화를 반영하여 새롭게 고찰해볼 필요가 있음

– 인터넷 경제의 규모를 보다 합리적으로 측정하는 과정에 유용한 시사점을 제공

○ 본 연구 주제인 인터넷 경제는 플랫폼 경제보다 개념적으로 큰 범위이며, 해외 사례의 경우에도 주장자마다 개념 및 범위가 상이하므로 관련 개념에 대한 학술적 연구가 필요

– 특히, 플랫폼 경제가 대표적인 양면 시장의 형태이며, 유형별로 모형을 통해 특징과 정책적 시사점을 분석할 필요가 있음

제 2 절 디지털 경제, 인터넷 경제, 플랫폼 경제의 개념과 특징

○ European Parliament(2015) 보고서에 따르면, ‘디지털 경제(digitalized economy)’는 통신 및 데이터 처리(communication and data procession) 등의 디지털 기술에 기반한 것으로서 ‘인터넷 경제’나 ‘온라인 경제’로 지칭되기도 함

– 공유경제나 IoT 등의 발전에서 볼 수 있듯이 디지털 경제는 물리적/오프라인 경제와의 연계성이 가속화되고 있으므로, 디지털 경제를 명확히 구획하는 것은 점점 더 어려워지고 있음

○ 디지털 경제를 위와 같이 정의한다면, 인터넷 경제는 좀 더 구체적으로 인터넷 통신에 기반한 경제를 의미하는 것으로 논리적으로는 디지털 경제의 한 부분을 의미한다고 볼 수 있으나, 사실상의 차이는 없는 것으로 볼 수 있음

– 지난 30년간 인터넷 보급 및 활용의 끊임없는 확대의 결과로, 오늘날의 주요 통신 및 데이터 처리 등의 디지털 기술 대부분은 인터넷에 기반하고 있기 때문

○ 이미 언급한 European Parliament(2015) 보고서에 따르면, 디지털 경제에는 여러 가지 특징들이 있는데 네트워크 효과(network effect)와 경합성(contestability)을 가장 대표적인 특징으로 주목할 수 있음

– 네트워크 효과란, 어떠한 상품/서비스를 이용하는 사람들의 수가 많아질수록 그 상품/서비스의 가치가 높아지는 현상을 말함

– 예를 들어, 우리나라에서는 모바일 메신저 서비스로 '카카오톡'을 가장 많은 사람들이 사용하는데, 카카오톡 서비스의 가치의 상당 부분은 바로 많은 사람들이 이용한다는 사실 자체로부터 기인함

○ 네트워크 효과는 시장의 집중화를 가져오는 경향이 있는 반면, 새로운 혁신적 기업들의 시장 진입에 의한 경합성은 시장의 구조적 변화를 초래할 잠재성이 있음

– 이러한 상반된 두 가지 효과가 공존함에 따라, 시장의 동태적 특성이 전통적인 산업들과는 근본적으로 다르게 나타남

○ 한편, 디지털 경제 또는 인터넷 경제의 논의에서 최근 큰 주목을 받는 또 한 가지 특징은 바로 플랫폼을 중심으로 생태계 및 가치사슬의 구조가 형성, 확장된다는 점

– 디지털 생태계 또는 인터넷 생태계에는 운영체제, 검색, SNS, 거래중개 등의 핵심기반 서비스들이 플랫폼으로의 역할을 하고 있음

– 이들 플랫폼은 직접 구체적인 콘텐츠나 서비스를 만들어 제공하기도 하지만, 그보다는 이용자와 판매자 또는 서비스 제공자, 이용자와 광고주, 이용자와 이용자 간의 상호작용을 돕는 작용을 주로 하면서 이익을 창출

○ ‘플랫폼 경제’는 광의의 의미로는 인터넷경제(디지털경제)를, 협의로는 인터넷 플랫폼을 통한 서비스 제공(카테고리 D)을 의미함

– 플랫폼 경제는 결국 디지털 경제나 인터넷 경제를 지칭한다고 볼 수 있으나, 플랫폼에 기반한 생태계 구조의 관점에서 바라본다는 시각을 포함한 개념으로 볼 수 있음

※ 플랫폼 경제가 인터넷 경제의 부분집합이므로 플랫폼 경제의 측정은 결국 인터넷 경제 측정과 동일함. 따라서 본 장에서는 플랫폼 경제의 개념과 특징을 살펴보고 플랫폼 경쟁의 유형에 따른 시장구조 특성을 제시함

○ 플랫폼 시장의 구조와 관련하여서는, 지난 10여 년간 경제학계에서 양면시장(two-sided market) 이론의 관점에서 상당한 학술적 연구가 이루어져왔으므로 이를 통해 플랫폼 경제를 이해하는 것이 디지털 경제 및 인터넷 경제의 구조적 특성을 이해하는 데 필수적이라고 볼 수 있음

제 3 절 플랫폼 경제 (양면시장) 이론

1. 양면시장 관련 개념과 사례

○ 양면시장(two-sided market)이란 두 그룹의 이용자(또는 사업자)가 특정한 플랫폼을 통해 상호작용하게 함으로써 새로운 가치를 창출하는 시장

○ 플랫폼은 시장 양측의 이용자들의 상호작용을 위한 기반 및 편의를 제공하고 시장을 형성하여 요금을 징수하는 역할을 하는 주체

– platform은 라틴어로 ‘평평한 형태의 그 무엇’인데 평평한 형태이기 때문에 이를 기반으로 무엇인가 만들어 올릴 수도 있고 무엇인가를 할 수 있음(플랫폼전문가그룹, 2012)

○ 결혼정보회사, 신용카드, PC운영체제, 미디어(방송, 신문, 인터넷 포털), 전화 착신접속 등이 양면시장의 대표적인 예

– 통신, 디지털 기술과 관련된 서비스들이 양면시장의 특성을 갖는 경우가 많음

○ 생산자와 소비자를 연결하는 대부분의 상업행위가 양면시장이 되는 것은 아님

– 플랫폼이 양측의 직접적 거래 또는 상호작용을 촉진하고 편의를 제공하는 경

우만 양면시장에 해당

- 예컨대, 쇼핑몰은 양면시장 플랫폼이지만 (이에 입주한) 개별 상점들은 양면시장 플랫폼으로 볼 수 없음

○ 양면시장이 성립하기 위해서는, 플랫폼을 통해 상호작용하는 서로 다른 그룹들이 존재해야 하고 플랫폼 부재시에 높은 거래비용이 전제되어야 함

- 플랫폼의 매개를 통하지 않고는 양측 그룹의 상호작용이 현실적, 경제적으로 어려운 경우에 양면시장이 성립

○ '인터넷'의 경우: 통신사업자의 망을 중심으로 온라인 사업자와 일반 이용자가 시장의 양측을 구성하여 상호작용하고 통신사업자에게 망 이용료를 지불하는 구조는 양면시장의 기본구조와 일치함

- 단, 온라인 사업자와 일반 이용자가 직접 자유롭게 거래하여 망 이용료 부담을 상호 재정산할 수 있다면, 통신사업자가 요금설정 기능을 제대로 할 수 없게 되므로 양면시장의 유효성이 사라짐
- 그러나, 정액제가 인터넷 요금의 주를 이루는 구조에서, 가입을 전제로 할 때 매몰/고정 비용에 해당하는 정액요금을 온라인 사업자와 일반 이용자가 상호 재정산하는 것은 불가능하므로, 인터넷은 양면시장의 조건을 만족하게 됨

2. 양면시장(플랫폼 경쟁) 이론

○ 양면시장 이론은 플랫폼이 어떻게 디지털 생태계를 활성화하고 유지해나가는지를 설명하는 새로운 경제이론

- 시장경쟁의 도입을 통해 경제의 성과를 높이는 것이 전통적인 경제학의 기본적 대명제라면, 양면시장 이론은 플랫폼의 조정 기능이 경제적 가치를 만들어내는 데 있어 경쟁보다 더 중요한 역할을 할 수 있음을 보여줌

○ 양면시장에서 플랫폼 사업자의 이윤극대화 전략의 핵심은, 양측 이용자들을 균형적으로 확보하여 플랫폼의 가치를 높이는 데에 있음

- 플랫폼은 양측 이용자들 간의 연결과 상호작용을 돕는데 그 과정에서 특히 플랫폼 이용료의 차별적 구조를 활용해 생태계를 만들어내고 자신의 수익도

확보함

○ 일반적인 시장에서 기업의 경쟁력은 가격과 품질에 주로 의존하는데, 양면시장에서는 이뿐 아니라 간접적 네트워크 외부성(indirect network externality) 효과를 창출하는 것이 중요하다는 특수성이 있음

– 간접적 네트워크 외부성이란, 어느 한쪽의 가입/이용 증가가 상대측의 효용을 발생시키는 현상을 말함

– 플랫폼 기업들은 간접적 네트워크 외부효과의 극대화를 추구하게 되며 이를 통해 양면시장 생태계의 가치가 창출

○ 특히, 요금 설정에 있어 어느 한쪽 그룹 이용자에게 상당히 낮은 (때로는, 원가 또는 한계비용보다 낮은) 요금을 제시하는 등, 기존의 교과서적 경제학 이론으로는 설명되지 않는 요금구조가 간접적 네트워크 외부성 극대화 전략의 수단으로 사용됨

– 요금수준뿐만 아니라 요금구조가 시장의 성과에 중요한 영향을 갖는 것이 양면시장의 특성

– 서로 다른 이용자 그룹 각각의 수요특성 및 그룹간 효과를 고려하여 양쪽 수요의 균형을 유지할 필요성

– 수요의 가격탄력성이 높아 요금할인에 의한 수요확대가 크게 나타나는 측, 그리고 상대측 이용자들에 파급되는 긍정적 네트워크 외부효과가 큰 측에 요금할인을 제공하는 것이 효과적

○ 예를 들어, 검색서비스 플랫폼은 검색 이용자에게는 무료로 서비스를 제공하지만 플랫폼의 또다른 이용자인 검색광고주들에게는 높은 광고비를 받음

– 무료로 서비스를 제공하기 때문에 많은 이용자들이 검색서비스를 이용하게 되고, 다른 한편으로 높은 광고수익에 기반하여 더 좋은 검색서비스를 제공함으로써 더 많은 검색 이용자들을 만들어낼 수 있음

○ 양면시장 플랫폼의 대표적 예인 신용카드 플랫폼은 그 가입회원에게도 회비를 받지만 대부분의 수입을 가맹점 수수료를 통해 획득함

– 카드 사용자들에게는 낮은 회비를 받을 뿐 아니라 다양한 혜택을 줌으로써 카드 사용을 유도할 수 있는데, 이는 충분한 가맹점 수수료 수익이 있기 때문에 가능

– 신용카드 사용 활성화에 따라 거래가 간편해지고 안전한 외상거래가 가능해짐으로써 수수료 비용을 부담하는 가맹점들도 결과적으로 혜택을 보게 됨

○ 이론적 연구에 따르면, 이러한 비대칭적 요금구조는 기업의 이윤극대화 전략이 될 뿐만 아니라 사회적 후생 관점에서도 효율적인 것으로 정당화됨

– 비대칭적 요금구조가 소위 chicken-and-egg 문제[6]를 해결함으로써 플랫폼이 활성화되고 간접적 네트워크 효과에 의한 가치가 새롭게 창출되기 때문임

○ 물론 이윤을 극대화하는 요금구조와 사회후생을 극대화하는 요금구조가 일반적으로 정확히 일치하지는 않음

– 양면시장의 균형에도 왜곡과 시장실패가 존재할 수 있으나, 양면시장 구조상에서의 시장 왜곡은 너무 복잡한 형태로 나타나기에 정책적으로 대응하기가 쉽지 않음(Monopolkommission, 2015)

3. 양면시장 이론의 경쟁정책적 시사점[7]

○ 양면시장에서의 효율적 가격구조는 원가구조에 비례하지 않음

– 원가정보에 기초한 최적요금 논의 및 규제정책은 오류의 가능성

○ 양면시장 한측에서의 높은 마진이 반드시 시장지배력을 의미하는 것은 아님

– 양측으로부터의 수익의 합이 플랫폼 비용을 지속적으로 현저히 초과하는지가 시장지배력 평가의 기준이 될 필요

○ 양면시장에서의 원가 이하의 가격 설정이 약탈적 가격을 의미하는 것은 아님

– 약탈 의도보다는 양면시장 플랫폼 가치/수익의 극대화 행위로 해석이 가능

○ 플랫폼간 경쟁의 심화가 반드시 효율적 가격구조로 이어지는 것은 아님

– 플랫폼 경쟁 증가로 가격구조가 더욱 왜곡되어 한 측이 다른 측에 비해 과도히 더 많은 비용을 부담하게 될 수 있음

– 이러한 배경에서 독일 독점위(Monopolkommission, 2015)는 플랫폼간 경쟁이

---

6) 플랫폼 사업자가 플랫폼에 참여하는 서로 다른 양 측면으로부터 참여에 대한 대가를 궁극적으로 지불하게 함에 있어 양 측 중 어느 쪽을 먼저 플랫폼에 끌어들여야 할 것인가를 설계할 때 발생하는 문제

7) 이 부분은 Wright(2004)의 내용을 주로 참고하였음

후생을 증대하는지의 여부가 명확하지 않다고 언급한 바 있음

○ 성숙된 시장에서도 비대칭적인 가격구조는 정당화될 수 있음

– 시장의 발전 초기의 chicken-and-egg 문제를 풀기 위해 비대칭적 가격구조가 허용된다고 볼 수 있으나, 이미 성숙된 시장에서도 간접적 네트워크 효과는 계속 존재하므로 비대칭적 가격구조가 여전히 정당화될 수 있음

○ 비대칭적 가격구조가 반드시 이용자간 상호보조를 의미하는 것은 아님

– 비대칭적 가격구조에 의해 플랫폼의 가치가 상승되는 측면까지 고려하면, 단순히 이용자간 제로섬 관계가 성립하는 것은 아님

○ 요금규제 정책에 대한 시사점

– 사업자의 유연한 요금구조 설정이 가능한 허용되는 것이 바람직

– 시장의 형성 초기에는 가입에 따른 네트워크 효과의 중요성이 높으므로 요금 관련 규제의 유연성이 더욱 요구

– 이용자의 가입/이용이 양면시장적 가치에 기여하는 바가 클 때 그에 상응하여 가입비/이용료가 낮게 책정되는 것이 타당하다는 점이 인식될 필요

○ 양면시장에서의 그룹간 상호의존성이 신규진입을 저해하는 진입장벽 효과에 주목할 필요

– chicken-and-egg 문제가 진입단계의 장애물로 작용

○ 이러한 양면시장적 진입장벽 효과는 특히 후발사업자에게 비대칭적 부담을 주는 문제

– 후발사업자 자신이 chicken-and-egg 문제를 해결해야 할뿐만 아니라, 이미 충분한 이용자 확보를 통해 높은 플랫폼 가치를 보유하고 있는 기존사업자와 경쟁해야 하는 부담이 있음

○ 시장의 집중이나 쏠림이 반드시 비효율성을 초래하는 것만은 아님에도 유의할 필요

– 경쟁이 감소하더라도 네트워크 외부성 증대에 의한 효율성이 발생하는 측면을 추가적으로 고려해야 함

– 이는 플랫폼간 합병승인 심사시 쟁점화될 가능성이 있음

○ 간접적 네트워크 외부성을 고려한 SSNIP 테스트를 수행하기 어려우므로 시장획정 및 그에 따른 시장지배력 평가가 어려워지는 문제가 있음

– 가설적 요금 인상시 시장 양측간 수요감소의 연쇄적 파급효과가 발생하므로 그 효과의 추정이 어려움
– 이러한 효과를 무시할 경우, 수요의 가격탄력성을 과소평가하여 시장을 지나치게 좁게 획정할 우려
– 경쟁적/효율적 가격수준을 사전에 파악하기 어렵기 때문에 SSNIP 테스트의 출발점인 기준가격 설정도 어려움

○ 이러한 경쟁정책 틀의 한계는 이미 예전부터 존재해왔던 것이나 플랫폼 시장의 발전에 따라 더욱 부각되게 된 것으로 볼 수 있음
– 과거에는 정책의 투명성, 일관성, 집행편의성(득)이 정책오류 가능성(실)보다 컸다고 볼 수 있음

○ 결론적으로, 시장지배력/경쟁제한성의 평가 및 그 대응에 있어서 규제기관의 능동적, 책임적인 역할이 요구
– 그러나 입법부와 사법부가 규제기관의 이러한 역할을 적절히 수용하는 환경이 되어야 한다는 근본적 문제가 존재

4. 플랫폼 경쟁의 유형에 따른 시장구조 특성

가. 싱글호밍과 멀티호밍에 따른 유형 구분

○ 이용자들이 플랫폼을 이용하는 행태가 싱글호밍(single-homing) 또는 멀티호밍(multi-homing)에 해당하는지 여부에 따라 양면시장 구조의 유형을 구분해 볼 수 있음
– 이러한 양면시장 유형의 구분은 곧 플랫폼 경쟁의 유형을 결정하는 기준이 될 수 있음

○ 싱글호밍이란 이용자들이 주로 하나의 플랫폼을 이용하는 경우를 말하고, 멀티호밍은 여러 개의 플랫폼을 동시에 이용하는 경우를 말함
– 예를 들어, 스마트폰 메신저는 여러 개의 서비스들을 동시에 이용할 수도 있지만 사용과정에서 번거로움이 발생하게 되어 대개의 경우 하나의 서비스만을 이용하는 싱글호밍의 경향을 보이게 됨

– 반면, 검색서비스나 인터넷 포털서비스는 굳이 하나의 서비스만을 이용할 필요성을 느끼지 못하므로 멀티호밍이 가능함

○ 양면시장 구조 또는 플랫폼 경쟁은 다음과 같이 3가지 유형으로 구분해볼 수 있음

– 첫째, 플랫폼의 양측 이용자가 모두 싱글호밍인 경우 (예: 결혼정보회사)

– 둘째, 플랫폼의 한측 이용자는 싱글호밍, 다른측 이용자는 멀티호밍인 경우 (예: 유료방송, 신문)

– 셋째, 플랫폼의 양측 이용자가 모두 멀티호밍인 경우 (예: 호텔예약 사이트, 인터넷 쇼핑몰)

○ 싱글호밍이 주된 형태로 나타나는 플랫폼은 당연히 더 쏠림현상이 쉽게 나타날 것으로 예상되므로, 플랫폼의 지배력은 멀티호밍이 증가할수록 약화될 것으로 보는 시각이 있음

– 특히, 싱글호밍 이용자에 대한 고착효과(lock-in effect)가 발생하면 이것이 플랫폼 지배력의 원천이 될 수 있음

– 이는 구글이 자신의 플랫폼 지배력이 크지 않음을 주장하기 위해 자주 사용하는 논리들 중의 하나임

○ 그러나 싱글호밍에 비해 멀티호밍이 항상 더 경쟁적인 시장환경을 만들어낸다고 할 수는 없음

– 싱글호밍의 경향을 보이는 이용자들은 하나의 플랫폼을 선택하기 원하는데, 이는 복수의 플랫폼들이 시장에 존재할 때 이용자의 선택을 받기 위해 서로 경쟁하는 관계에 있음을 의미

– 반면, 이용자들이 멀티호밍의 경향을 보이는 경우에는 여러 플랫폼을 함께 보완적으로 이용하고 있음을 시사하므로, 이들 플랫폼은 실질적으로 서로 강하게 경쟁하는 관계에 있지 않을 수 있음

– 따라서, 멀티호밍 환경에서는 여러개의 플랫폼들이 있다 하더라도 오히려 이들이 각각 독자적 지배력을 발휘하게 될 가능성이 있음

나. 싱글호밍－싱글호밍의 경우

○ 플랫폼이 상호 배타적인 연결성(접근)을 제공하는 가장 단순한 형태의 양면시장

－ 플랫폼간 경쟁의 양상도 비교적 명확

○ 네트워크 효과 발생이 제한적이므로 플랫폼이 이러한 구조를 추구할 유인은 크지 않으나, 선발 플랫폼이 비호환성 전략을 통해 진입장벽을 구축하려고 할 경우에 활용될 수 있음

○ 이러한 플랫폼 구조에서는 플랫폼 경쟁이 진행됨에 따라 네트워크 효과가 쏠림현상을 유발하는 문제가 있음

－ 간접적 네트워크 효과가 직접적 네트워크 효과처럼 작용 (이에 따라 양면시장 비대칭 구조의 역할은 감소)

－ 지배적 플랫폼이 출현하게 되면 (잠재적) 경쟁/진입이 감소하게 되고, 지배적 플랫폼은 투자전략 단계에서 수확전략으로 전환하고 혁신유인은 감소하게 됨

－ 물론 플랫폼들의 차별성이 유지된다면 차별화된 플랫폼들 간의 지속적 경쟁이 가능할 수도 있음

○ 이러한 구조에서 지배적 플랫폼에 대한 규제는 전통적인 경쟁정책, 즉 경쟁을 활성화하고 보호하는 정책에 초점을 둘 수 있음

－ 그러나, 서로 다른 플랫폼들 간의 비호환성이 유지됨에 따라 네트워크 효과의 효율성이 저해되는 문제점을 함께 고려할 필요

－ 지배적 플랫폼으로의 쏠림과 그 효율성을 인정하고, 규제산업을 전제로 한 소비자 이익 보호를 위한 (요금/품질) 규제에 초점을 두는 것이 대안적인 정책이 될 수도 있음

다. 싱글호밍－멀티호밍(경쟁적 병목)의 경우

○ 플랫폼을 이용하는 한쪽 그룹은 싱글호밍, 다른 한쪽 그룹은 멀티호밍인 경우는 양면시장 모형 중에서 특별히 많은 정책적 관심의 대상이 되는 유형으로서 ‘경쟁적 병목(competitive bottleneck)’이라고 불림

○ 예를 들어, 일간신문 구독자는 대개 하나의 신문을 선택하는 싱글호밍 이용자인 반면, 광고주는 여러 종류의 신문과 매체에 광고를 게재하는 멀티호밍 이용자에 해당하므로, 신문과 같은 광고기반의 미디어들은 대개 경쟁적 병목 모형으로 설명되는 유형의 양면시장에 해당한다고 볼 수 있음

○ 경쟁적 병목 구조에서 플랫폼은 경쟁을 통해 확보한 싱글호밍 이용자들에 대한 접근권을 독점함으로써 초과이윤을 획득할 수 있음

– 플랫폼이 싱글호밍 이용자들을 확보하게 되면 이에 따라 플랫폼들은 (반대측 이용자 입장에서) 완전히 차별화되는 효과가 있으며, 반대측 이용자들은 이들 플랫폼을 함께 이용하려는 멀티호밍의 유인을 갖게됨

– 플랫폼은 멀티호밍 이용자에 대해 강력한 지배력을 행사할 수 있게 됨

– 싱글호밍측에 대해 치열하게 서로 '경쟁'하던 플랫폼들이 그 반대쪽의 멀티호밍 이용자 시장에서는 경쟁이 사라지고 각각 독점적 지배력을 갖는 '병목'의 지위를 갖게 되므로, 이러한 구조가 '경쟁적 병목'이라는 이름을 갖게 되는 것임

○ 예컨대, 케이블TV나 IPTV와 같은 유료방송 플랫폼들은 유료방송 가입자(싱글호밍)들을 대상으로 치열하게 요금할인 경쟁을 벌이지만, 플랫폼의 반대측에 있는 홈쇼핑방송사업자(멀티호밍)들로부터는 높은 홈쇼핑송출수수료 수익을 벌어들임

– 인터넷과 스마트폰 생태계에서는 이와 같은 정도로 명확히 경쟁적 병목의 구조를 갖지 않는 경우들도 많지만, 어느 한측에서는 치열히 경쟁하면서 그로부터 확보된 가입자를 기반으로 다른 한측에서 높은 수익을 올리는 경우를 흔히 찾아볼 수 있음

– 이는 경쟁적 병목 이론이 디지털 생태계에서도 유용한 레퍼런스가 될 수 있음을 의미

○ 일견 멀티호밍 이용자는 플랫폼 선택권을 보유한 것처럼 보이지만, 실제로는 플랫폼이 이미 확보한 싱글호밍 이용자들에게 접근하기 위해서는 플랫폼에게 이용료를 지불하는 것 이외에 대안이 없게 됨

– 이러한 이유에서 Armstrong(2006)은, 경쟁적 병목 시장에서는 멀티호밍 이용자들이 플랫폼들을 비교하여 선택하는 입장에 있지 않고 특정한 플랫폼의 선택은 다른 플랫폼 선택과 독립적으로 이루어지므로, 멀티호밍 이용자들을 확보하기 위한 플랫폼간 경쟁이 전혀 존재하지 않는다고 보았음

○ Armstrong은 경쟁적 병목 시장의 특징으로 다음 3가지를 지적하였음

– 첫째, 플랫폼은 싱글호밍 이용자들에 대한 접근권을 멀티호밍 측에 제공하는 과정에서 독자적 지배력을 가짐

– 둘째, 멀티호밍측에 대해 지나치게 높은 요금이 책정되므로 멀티호밍 이용은 사회적 적정수준보다 위축됨

– 셋째, 멀티호밍측에서 얻어진 플랫폼 초과이윤의 상당 부분은 싱글호밍 이용자들의 혜택으로 전환됨

제 5 장 결 론

1. 주요 연구 내용 및 결과

○ 인터넷이 경제전반에 미치는 영향력이 확대됨에 따라 인터넷경제의 규모와 경제적 파급효과를 추정하기 위한 방법론 구축에 대한 필요성이 확대되고 있음

○ 이러한 시점에서 본 연구는 네덜란드(CBS)의 인터넷 경제 측정 방법론을 우리나라 통계청의 공공데이터와 민간데이터에 적용하여 시산하고, 네덜란드 방법론에 대한 한계점 및 대안을 검토하여 우리나라 상황에 맞는 인터넷 경제 측정 방법을 제시하기 위해 아래와 같은 내용을 도출했음

- 첫째, 인터넷 경제, 플랫폼 경제 문헌 연구를 위해 인터넷 경제의 개념, 포괄범위 등에 대한 국내외 다양한 문헌연구 및 해외 사례를 연구함
- 둘째, 네덜란드 통계청의 인터넷 경제 측정 방법론 적용 연구를 위해 네덜란드 통계청이 활용한 빅데이터 기법을 우리나라 데이터에 적용하고 시사점을 제시함
- 셋째, 플랫폼 경제 측정방법론 연구를 위해 4차 산업혁명의 기반이 되는 플랫폼 경제의 개념과 특징을 살펴보고 플랫폼 경쟁의 유형에 따른 시장구조 특성을 제시함

○ 본 연구의 핵심 부분인 네덜란드(CBS)의 인터넷 경제 측정 방법론을 우리나라 통계청 데이터와 다양한 공공, 민간데이터에 적용하여 시산한 결과는 다음과 같음

- (데이터) 데이터프로바이더(DP)가 크롤링한 데이터와 이를 보안하기 위한 외부 자료(공정위 “통신판매사업자 신고자료”, 통계청 “온라인쇼핑동향조사 모집단 자료”, “KED 웹사이트자료”)를 활용하여 통계청의 기업등록부(BR)와 매칭

- (분포) 인터넷 경제 분석에 활용한 최종 데이터는 기업체 기준으로 총 216,099개이며, 그룹별 분포를 살펴보면 B그룹(웹사이트 보유) 46%(99,391개), C그룹(온라인 상점) 53% (114,619개), D그룹(온라인 서비스) 0.9%(1,916개), E그룹(인터넷관련 ICT기업) 0.1%(173개)로 나타났음
- (매출액) 인터넷 경제에 속한 기업체의 총 매출액은 2,837조 원이며, B그룹이 70.9%(2,011조 원), C그룹이 28.6%(811조 원), D그룹이 0.4%(13조 원), E그룹이 0.1%(2조 원)을 차지하는 것으로 나타남
- (종사자) 인터넷 경제에 속한 기업체의 총 종사자 수는 6,467,073명이며, B그룹이 70.9%(4,580,103명), C그룹이 28.3%(1,832,580명), D그룹이 0.7%(46,359명), E그룹이 0.1%(8,031명)을 차지하는 것으로 나타났음
- (지역별 분포) 기업등록부내 인터넷 경제 기업의 경우 서울(36.8%), 경기도(26.3%), 부산(5.0%) 순으로 분포가 높았는데 인터넷 기업이 서울과 경기 등 수도권에 밀집된 정도가 상대적으로 높은 것을 볼 수 있음
- (산업별 분포) 기업등록부내 인터넷 경제 기업의 경우 도매 및 소매업(34.3%), 제조업(21.4%), 출판, 영상, 방송통신 및 정보서비스업(8.1%) 순으로 분포가 높은데 전체 기업대비 도매 및 소매업의 비중이 높으며, 특히 제조업의 비중이 높은 것이 특징

2. 연구의 한계 및 시사점

가. 연구의 한계

○ (구매 데이터) 인터넷 웹크롤링 데이터의 품질 미흡으로, 웹사이트를 보유·운영하는 사업주체 파악이 곤란. 특히 연계키로 활용되는 사업자등록번호 정보가 부실함

○ (웹크롤링업체) 데이터프로바이더는 웹크롤링 전문 업체이나, 한글 웹페이지 크롤링은 전문성이 부족

○ (연계키의 품질) 전화번호의 경우 번호체계를 활용, 식별부호－국－번호의 구분이 필요하나 자릿수 부적절, 이상치(0000 등)가 상당수 발견

– 또한 관련 세무법인 전화번호를 해당 웹사이트에서 제공하는 경우가 다수발견되어 동일한 전화번호가 3회 이상 중복되는 경우 연계에서 제외하였으며, 사업체명, 주소의 경우도 표준화 부재 등으로 연계키로 활용할 수 없었음

○ (매칭 데이터) 통계청 BR/온라인쇼핑몰, KED, 공정위 등 다양한 소스를 활용하였으나 최종 매칭 데이터의 포괄 범위(coverage)에 대한 신뢰도 확보 곤란으로 산출된 기초 통계 결과도 제한적 해석만 가능

○ (매칭 방법론) 카테고리 분류(특히 D, E)시 활용한 키워드 판정 방법론의 신뢰도 검증이 필요하며, 동시에 여러 속성을 지닌 업체를 한 업종으로 강제 할당하는 방법에 대한 검토가 필요함

– 이는 3장에서도 기술하였듯이 온라인과 오프라인을 통해 다양한 서비스를 제공하는 은행 등과 같은 기존 기업체의 경우 하나의 카테고리로 분류하고 있으나 매출이나 인력 등의 비중에 관한 보조정보가 없는 상황에서는 강제 할당에 대한 객관적 근거가 불충분

나. 시사점

○ 본 연구는 인터넷경제 규모 측정을 위해 빅데이터 방법(크롤링)을 통해 수집한 웹사이트 데이터와 통계청의 기존 산업통계를 결합하는 방법을 활용한 파일럿 성격의 연구임

– 크롤링 데이터의 품질, 매칭을 위한 연계키, 인터넷 경제 카테고리 분류를 위한 보조 정보의 부재, 매칭 방법론의 정합성 등의 측면에서 판단할 때, 전체 경제에서 차지하는 인터넷 경제의 규모를 산출하기에는 시기상조인 것으로 판단함

– 신뢰할 수 있는 인터넷 경제 규모 측정 결과를 산출하기 위해서는 웹사이트 모집단 정보의 확보, 웹크롤링 결과의 품질 제고, 해당 기업체 단위의 온라인 및 오프라인 매출 비중을 추정하기 위한 보조 정보의 확보 등이 필수적이며 다음과 같은 추정을 위한 여건과 관련 기초 연구가 향후 필요할 것으로 판단됨

○ (웹사이트 모집단 정보) 인터넷 경제 측정의 출발은 웹사이트 모집단 정보 확보임

– 대안 1: 통계청의 주요 조사(전국사업체조사, 광업제조업조사 등) 시 사업체(기업체)별 웹사이트 조사

※ 웹사이트 주소를 조사항목으로 포함하고 있는 기존 조사(예: 방송산업실태조사 등)를 확인 후 이를 활용하는 방법 검토 필요

– 대안 2: 한국인터넷진흥원 도메인 등록시 필요 정보(업체명, 사업자등록번호 등) 기입 의무화 방안. 단 기존 등록 정보는 결측 항목 조사 등 정비가 필요(단점은 .kr 이외 .com 등 도메인 정보는 확보 곤란)

※ 인터넷 사이트 기준의 승인통계(정보접근성실태조사) 신뢰도 확보 위해서도 필요

○ (웹크롤링) 통계청 BR데이터와 매칭 성공률을 높이기 위해서는 필요 정보(사업체명, 사업자 등록번호, 전화번호, 쇼핑카트 유무, 취급 품목 등) 수집을 위한 크롤링의 성능이 매우 중요하므로 기술력이 확보된 국내 업체와 협업 또는 자체적으로 크롤링 능력 확보가 필수

– 기존 통계나 행정자료가 포괄하지 못하는 기업 특성을 보완하는 것이 크롤링 등 빅데이터 방법을 사용하는 주된 이유임을 감안할 때, 기존 자료에서 미흡한 영역의 정보를 확보하기 위한 방법론 개발이 중요

– 한국어로 구성된 홈페이지 분석 능력이 중요하므로, 웹크롤링에 대한 충분한 기술력을 보유한 국내 업체를 통해 연구 수행 가능성을 모색할 필요가 있음

○ (인터넷경제 카테고리 분류) 크롤링 데이터를 통해 카테고리를 자동 분류하는 판별 모형의 성능이 매우 중요하므로 이를 위한 빅데이터 방법론 개발이 필수

〈표 5－1〉 인터넷경제 카테고리 분류별 판정 난이도

구분	판단 기준	난이도
B	웹사이트 유무	낮음
C	쇼핑 카트 유무	보통
D	특정 취급 품목(서비스) 유무	매우 높음
E	특정 취급 품목(서비스) 유무	높음

○ (기초 연구) 카테고리 분류(특히 D, E)시 활용한 키워드 판정 방법론의 성능 검증, 인터넷 경제 규모 추정 결과 신뢰도 측정 등 관련 기초 연구 필요

○ (기업 정보) 인터넷경제의 특성을 고려하여 정확한 경제 규모를 산출하기 위해서는, 기업의 매출액 등 주요 지표를 온라인 기반과 오프라인 기반으로 나눌 수 있는 보조 자료가 필요

– 기업등록부 등 행정자료에서는 기업의 총 매출액에 대한 정보만 제공되고 있음을 고려할 때, 이를 보완할 수 있는 자료 구축이 필요

※ 웹사이트를 보유하였으나 온라인 매출 규모가 미미한 기업의 총 매출액을 인터넷경제 규모 추정에 포함할 경우 과대추정 발생

○ (한국형 측정 방법) 본 보고서의 분석은 네덜란드 통계청의 방법론을 기반으로 한 것이므로, 한국의 산업 특성을 반영한 인터넷경제 측정 방법을 발전시켜 나가야 함

– 한국의 인터넷경제의 특성을 다른 나라와의 국제비교를 통해, 한국 특성에 부합하는 고유 방법론을 구축하는 것이 필요

– 오픈마켓 등 인터넷경제에서 중요한 대상을 포괄할 수 있는 측정 방법 필요

– 인터넷경제를 파악할 때 전통적인 산업 분석에서는 고려하지 않았던 지표 및 특성이 중요하므로, 다양한 관점의 통계지표 개발을 위한 접근이 요구됨

※ 예를 들어 인터넷 경제 측정에 있어서, 전통적인 사업체의 지리적 소재지 분포는 큰 의미를 갖지 못할 가능성이 있음

다. 기대효과

○ (학술적 효과) 거시 및 산업 데이터인 국민계정(SNA) 데이터와 산업분류를 이용한 기존의 학술 연구와 달리 네덜란드 측정방법론은 미시데이터를 연계하는 방식과 빅데이터 방법론을 활용한 방법론의 제시를 통해 학술적 측면의 새로운 방법론 적용에 따른 연구 진작과 다양한 시사점을 도출할 것으로 보임

○ (정책 효과) 파일럿 프로젝트 과정에서 도출된 다양한 문제점을 기초로 정확한 인터넷 경제 규모 측정을 위해 필요한 모집단 정보 확보, 온・오프라인 매출 분리를 위한 보조 정보의 확보 등에 대한 필요성 제시를 통해 향후 관련 정책 지원을 위한 기초 데이터 정비 방안 수립에 기여

참 고 문 헌

〔국내 문헌〕

미래창조과학부, “전기통신번호관리세칙”(미래창조과학부고시 제2015-52호, 2015. 7. 27., 일부개정), (http://www.law.go.kr/LSW/admRulLsInfoP.do?admRulSeq=2200000038186)

이경선 외(2012. 11),『인터넷의 경제적 파급효과 분석』, 방통융합미래전략체계 연구 지정2012-10, 방송통신위원회.

정용찬(2016), 빅데이터의 국가통계 활용 전략. 통계연구 특별호 1-16. 통계청 통계개발원.

정용찬・주재욱・이원태・김윤화・유선실・김욱준・오윤석・박민규・황용석・황선웅(2016), 조사환경 변화에 대응한 ICT 통계 생산체계 혁신방안 연구(I) 총괄보고서. 정보통신정책연구원.

정용찬・이원태・정혁・김윤화・유선실・정부연・오윤석・박민규・권헌영・오형나(2017), 조사환경 변화에 대응한 ICT 통계 생산체계 혁신방안 연구(II) 총괄보고서. 정보통신정책연구원.

통계청(2018), “통계빅데이터센터 온라인서비스 및 자료연계시스템 개발 ISP 수립”, 중간보고 자료.

플랫폼전문가그룹(2012),『플랫폼을 말하다』, 클라우드북스.

〔해외문헌〕

Armstrong, M.(2006). Competition in Two-sided Markets, RAND Journal of Economics, Vol.37.

Boston Consulting Group(2010).『The Connected Kingdom: How the Internet is Transforming the U.K. Economy』.

Boston Consulting Group(BCG)(2011). 『Turning Local: From Madrid to Moscow, the Internet is going Native』.

_____(2012). 『The $4.2 Trillion Opportunity: The Internet Economy in the G-20』.

CBS(2016). Measuring the internet economy in The Netherlands: a big data analysis.

Deloitte(2011). "The Connected Continent: How the internet is transforming the Austrailian economy." 2011. 8.

European Parliament(2016). "Challenges for Competition Policy in a Digitalised Economy".

Ilsoon Shin, Jungsoo Park, Seonghoon Jeon(2012). "The Size, Effects, and Efficiency of Internet Economy in Korea", International Telecommunications Policy Review, Vol. 19, No. 2, 2012.

McKinsey Global Institute(2011). "Internet matters: The Net's sweeping impact on growth, jobs, and prosperity". 2011. 5.

McKinsey Global Institute(2011. 7). 『The Impact of Internet technologies: Search』.

Monopolkommission(2015). "Competition policy: The challenge of digital markets," Special Report No 68.

OECD(2008). "OECD Information Technology Outlook 2008".

_____(2011). "OECD Communications Outlook 2011".

_____(2012). "OECD Internet Economy Outlook 2012".

_____(2013). "Measuring the Internet Economy: A Contribution to the Research Agenda", OECD Digital Economy Papers, No. 226, OECD Publishing.

_____(2015). "OECD Digital Economy Outlook 2015". OECD Publishing.

_____(2017). "OECD Digital Economy Outlook 2017".

Statistics Netherlands(2016). "Measuring the internet economy in The Netherlands: A big data analysis", CBS Discussion Paper, 2016.

Wright, J.(2004). One-sided Logic in Two-sided Markets, Review of Network Economics, Vol.3, Issue 1.

〔통계 사이트〕

공정거래위원회 통신판매사업자(http://www.ftc.go.kr/www/dataOpen.do?key=259)

NICE평가정보원 kisvalue DB(https://www.kisvalue.com/web/index.jsp)

데이터프로바이더 DB(www.dataprovider.com)

통계청(http://kosis.kr/index/index.do)

한국기업데이터(KED) 인터넷주소 보유 기업 리스트 DB

〔부 록〕 데이터 프로바이더 제공 데이터의 변수 설명

섹션	변수	항목수	내용	정확도	범위
Geolocation			Our geolocation filters give an indication of the real-world geographic location of a website or the business behind the website, for example in which country a website is based or what zip code it is connected to.		
	Continent	928,218	Continent refers to the geographical area where the website is located or where the website originated from. We determine the continent of a website based on where it is hosted as well as the top-level domain that is used. For example: North America.		100%
	Country	928,218	Country refers to the land where the website is located or where the website originated from. We determine the country of a website based on multiple variables such as hosting country, language on the website, top-level domain and contact details.	90%	100%
	Zip code	34,456	Zip code refers to the postal code that is found on a website. If our spider comes across more than one zip code on a website, it uses the code found on the contact page. This zip code also helps us determine the country of a website: our spider uses zip code information as a confirmation of country.	96%	4%
	Zip code quality	32,954	Zip code certainty refers to the expected quality of a zip code found on a website, and the certainty that it is correct. It ranges from 0 (uncertain) to 100 (certain). A zip code found on the contact page, for example, has a high certainty of being correct. A zip code on the Terms on Conditions page on the other hand, has a lower certainty.		4%
	City	19,678	City refers to the location of a website. We do, however, not measure this field directly. We have a database with zip codes from all over the world and know which zip code is connected to which city. This way we base the city shown on the zip code that our spider finds on the website. For example: San Francisco.	96%	2%
	Address	2,297	A company's address can be difficult to determine based on the many forms it may take. We thus base our address information on the company registration at the Chamber of Commerce. If a company states a Chamber of Commerce number, our spider will crawl and use this number. The address connected to the number will be shown as the company's address.	96%	0%
	Addresses	76	Addresses refers to all secondary addresses a company provides on its website. Please note that the most prominent address is already shown in the address column: this field only contains any additional addresses found. For example: 610 West 5Th Street Wilton; 52778,10020 West Us Highway 150; Edwards, 61528.	96%	0%
Business			Businesses often use a website to promote their services or products. If a website is connected to a business, this segment will provide you with data on the company type, payment information and contact details.		
	Company name	146,331	Company name refers to the name of a company that owns the website. We use proprietary techniques to retrieve the name of the company. If no name can be found on the website, this field will stay empty.	90%	16%
	Company type	14,979	Company type refers whether the organization is of the non-profit or for profit kind. We determine this information based on the Chamber of Commerce number, or the company name that we find on a website.	98%	2%
	Company quality	146,331	Company certainty refers to the expected quality of a company name found on the website and the certainty of it being correct. It ranges from 0 (uncertain) to 100 (certain). If a company name is found on the contact page, for example, it has a high certainty. A company name found in the Terms of Conditions has a low certainty.		16%

섹션	변수	항목수	내용	정확도	범위
	Legal entity	14,345	Legal entity can refer to the following: an association, individual, a corporation, trust, partnership or proprietorship that has legal rights to enter into contracts, agreements and such. The legal entity is extracted from the legal company name. For example: INC.	94%	2%
	Business Registry number	13,625	This refers to the number that is given to a company after registration with the Chamber of Commerce or sort like institution. We locate this number and verify the company by finding it in the Chamber of Commerce database. The Chamber of Commerce is not available in each country.	93%	1%
	Bank account number	32	A company's bank account number is taken from the website. We analyze the website and look for the specific format of a bank account number for this. If we find one, we will also try and determine an IBAN and BIC number. This is not available in each country.		0%
	IBAN number	32	The International Bank Account Number (IBAN) is an internationally agreed system for identifying bank accounts across national borders to facilitate the communication and processing of cross-border transactions. We analyze a company's website and look for the IBAN based on its format.	97%	0%
	BIC number	24	BIC refers to the standard format of Bank Identifier Codes (SWIFT) and is the unique identification code of a bank. These codes are used when transferring money between banks, particularly for international wire transfers. We analyze the website and look for the BIC number based on its format.	98%	0%
	Tax number	11	Tax number refers to the tax number found on a company's website. Our spider is trained to recognize the tax numbers from different countries based on their format, and can thus collect different kinds of numbers. This is not available in each country.	100%	0%
	Phone number	273,920	Phone number refers to the primary phone number on a website. Our spider starts out by looking for a phone number on the home page, and carries out the same procedure for all other indexed pages. The phone number that is mentioned the most is considered the primary phone number.	97%	30%
	Phone number certainty	273,920	Phone number certainty refers to the expected quality of a phone number found on a website and the certainty of it being correct. It ranges from 0 (uncertain) to 100 (certain). If a phone number is found on the contact page for example, it has high certainty. A number found on the Terms and Conditions page has a low certainty.		30%
	Secondary phone numbers	153,383	Phone numbers refers to all other phone numbers found on the website which are not considered as the primary number(s). We find all other phone numbers by going through all the pages of a website.	97%	17%
	Email address	32,959	Email address refers to the main email address a company provides on its website. This one is selected as the main address because it is used on classified pages such as the contact page, or because it is repeated multiple times. This field can only contain general email addresses, not personal ones.	96%	4%
	Secondary email addresses	138,943	Secondary email addresses are all additional email addresses that a company provides on its website. The address determined as the main email address is not included in this field. This field can contain both generic and personal email addresses.	95%	15%
	Contactable	334,246	Contactable indicates the ways in which the owner of a website can be contacted. Most websites contain several ways to contact the owner. Examples of possible ways of contact are phone number, email address, post address or social media.		36%
	Title	677,602	Title refers to the title found on the homepage of a website, often in the HTML of the page. If there is no available title on the homepage of a website, our spider proceeds to check page after page until it finds a title that it can display here.		73%

섹션	변수	항목수	내용	정확도	범위
	Description	419,621	This is the description written on a website, it gives an indication of what the website is about. A description can usually be found on the homepage, but if nothing is found there our spider will continue to crawl other pages until it finds one. If no description is found at all, we generate one based on the first paragraph of text on the homepage.		45%
	Keywords	752,066	Keywords refer to words that are frequently mentioned on the website. We generate these keywords by going through all pages and content on a website, and looking for words that are mentioned with high frequency. A bike rental website for example, will probably generate the keywords bike, hire and rent.		81%
	Relevant keywords	135,896	This field refers to the most relevant keywords that are found on a website and describe its content. Through language processing techniques our spider knows how to detect the most important words in a sentence and display these keywords here for an accurate reflection of the subject and contents of a website.		15%
	Word count	928,195	This word count gives an indication of the average number of words found on all indexed pages, based on the textual content of a website. For example: if a homepage contains 300 words and the contact page 200 words, the word count shown will be 250 words.		100%
	Website type	928,195	There are many kinds of websites. Possible website types are: blog, forum, business, eCommerce, placeholder (default website provided by a registrar), parking (a domain for sale, reserved or with just ads), empty, framed (one-pager that contains an iFrame displaying content of another website) or disallowed (with robot.txt).		80%
	Category	103,627	Category indicates the content topic. We take all the keywords from the homepage and use a bayesian algorithm with keywords associated with each categories for multiple languages to determine the best possible category. If there is not enough content on the website we can't identify a category.	66%	11%
	SIC code	37,058	The Standard Industrial Classification (SIC) is a system for classifying industies by a four-digit code. Dataprovider.com uses machine learning to find the best SIC code based on the textual content of a website. These SIC codes can be grouped into progressively broader industry classifications: major group and division.		4%
	SIC major group	37,058	The Standard Industrial Classification (SIC) is a system for classifying industries by a four-digit code. Dataprovider.com uses machine learning to find the best SIC code based on the content of a website. The major group is the broadest definition of an industry.		4%
	SIC division	37,058	The Standard Industrial Classification (SIC) is a system for classifying industries by a four-digit code. Dataprovider.com uses machine learning to find the best SIC code based on the content of a website. This division is the overarching category.		4%
	Copyright	159,769	Copyright refers to the date that a website was published or that it was last maintained. Our spider determines the copyright year by using the copyright field in the footer of a website. If no copyright date can be found, this field is left blank.	90%	17%
	Logo	179,622	Logo refers to the image URL that is most likely the logo of the website. This will often be found in the header or footer. We only detect the logo if the image is clearly visible in the HTML code. Logo's generated with Flash, JavaScript or CSS cannot be detected.	95%	19%
	Language	514,917	We currently recognize languages from 50 different countries. Our spider determines the language of a website by using N-Gram language detection. We then use this language to index the content of the website and confirm the country that a website is based in.	93%	55%
	Multi language	928,195	Multiple language refers to whether the website offers more than one language option. Our spider determines this by scanning the website and looking for the option of a language change. These options are usually found at the top of the front page.	37%	100%

섹션	변수	항목수	내용	정확도	범위
	Authors	111,736	Author refers to the META Author Tag that is used on the website. This tag shows the author of the website, and can also contain multiple authors. If no author can be found, this field will be left blank.		12%
	Privacy sensitive	928,215	Privacy sensitive indicates that the website stores personal information such as contact- of payment details. Websites that offer the possibility to create an account, log into a system, fill in a contact form or process a payment are flagged as privacy sensitive.		100%
	Pages	928,195	Pages refers to the estimated amount of pages a website contains. It is accurate up to 500 pages, which means larger websites are indexed at 500 pages. For our spider to be efficient, we only index 10 – 20 pages and review the internal links to provide an estimation of the amount of pages a website contains.		100%
	Pages indexed	928,195	To keep our spider efficient and not overtax websites, we do not download all the pages a website contains. Pages indexed refers to the actual number of pages of a website that is indexed and used for additional information.		100%
	Page types	620,836	To keep our spider efficient, we do not download all pages of a website. Pages such as contact, legal or FAQ contain the most important information and are thus prioritized during an indexation. Page types contains a list of all classified pages we have indexed.		67%
	HTML size (kb)	923,643	HTML size refers to the size of the homepage in KB (kilobytes). The size of the homepage is an indication of the size of other pages on the website. A small homepage usually indicates a website with little content, whereas a large homepage indicates more content.		100%
	Date found	928,218	Date found refers to the date that our spider has first encountered the website. It is possible that the website existed before this time. We crawl the web on a monthly basis, so only update our database per month.		100%
	Date analyzed	928,218	Date analyzed refers to the last time our spider has analyzed the website. All the data available from the websites was thus retrieved on this date. We crawl the web on a monthly basis, so this date may be a month in the past.	100%	0%
Ecommerce			We use a machine learning system to determine whether a website is an online store. Once we recognize a website as online store, this segment shows you the details of that store. This includes among other things shopping cart system, payment methods, payment providers and currency.		
	Online store	928,195	Online store indicates whether a website contains an online store. We have used machine learning classifying thousands of websites as eCommerce or non-eCommerce to help our spider determine if a website is an online store. This field not just refers to online stores, but also stores where you can purchase online and pick up later.	79%	100%
	eCommerce certainty	928,195	We have used machine learning to train our spider in recognizing eCommerce websites. This certainty score ranges from 0 (uncertain) to 100 (certain) and is an indication of how sure the spider is that it is indeed an online store that has been indexed.		100%
	Shopping cart system	10,170	Shopping cart system is the type of software used on eCommerce websites, that allows customers to place products in their cart and proceed to a checkout. We currently support over 100 different shopping cart systems, such as PayPal and Magento.	86%	1%
	Trustmarks	846	Trustmarks notify customers that a website is safe to purchase from by a logo placed on the homepage of a website. Our spider recognizes all Trustmark logos based on file name, title and alt texts. We currently do not recognize logos through image recognition.	94%	0%
	Delivery services	1,839	Delivery services refer to the ways in which eCommerce websites send out goods to their customers. Detection of these services is based on delivery logos, our spider crawls the website for specific pages with information on how packages are being shipped.	67%	0%

섹션	변수	항목수	내용	정확도	범위
	Payment methods	3,766	Payment methods are the systems that enable electronic payments on eCommerce websites. Our spider indexes payment methods based on the logos it finds on the website. If it is unable to find these logos, it tries to find a separate page that describes which payment methods are available.	87%	0%
	Payment providers	4,496	Payment service providers refer to the gateway provided to complete the actual transaction. Paypal and IDeal are examples of both a payment method and a payment gateway. Detection is done via logo detection and by looking for specific keywords in the payment options page.	63%	0%
	Currency	109,894	If a currency is mentioned on the website, this field shows the currency being used. We show the standard currency abbreviation such as USD or EUR. When multiple currencies are available, this field shows the currency that corresponds with the region the online store is based in.	94%	12%
Marketing			Our marketing filters provide information on the economic impact and size of a website. In addition, it shows you the marketing tools that are used by a website such as social media types, ad network and analytics.		
	Economic footprint	928,195	Our Economic Footprint is a proprietary rank that gives an accurate estimation of the economic impact of a website. This score can be used as a proxy for the number of visitors, the turnover and the size of a company. The scale runs from 0 (small) and 100 (big).		100%
	Economic footprint delta	236,255	Our Economic Footprint is a proprietary rank that gives an accurate estimation of the economic impact of a website. The delta value of the Economic Footprint is the difference between now and last year. The delta indicates if a site is growing or declining. This score can be used as a proxy for growing companies.		25%
	Heartbeat	513,650	Every month we visit the website and we compare the results with the previous months. Depending on the number of changes (and what has changed) we calculate a score, we call this score the Heartbeat. Heatbeat can be flat, very low, low, medium, high, very high or extreme.		55%
	Changes	909,049	We re-index all websites in our database each month. Changes keeps track of all the changes that have occurred and shows the number of changes since the last indexation. For more information on what has changed, visit a website via our Lookup tool.		98%
	Incoming links	928,164	Incoming links provides the number of hyperlinks that have been found directed towards the indexed website, coming from other websites. These incoming links are all unique. If a link to www.dataprovider.com is found 5 times on 1 website for example, it counts as 1 unique link.		100%
	Outgoing links	928,195	Outgoing links provides the number of hyperlinks found on the indexed website that point to a destination outside of the website. The number of outgoing links indicates the number of unique external links. If Dataprovider.com links to another website 5 times for example, this is counted as 1 unique link.		100%
	Site traffic estimation	115,333	This field shows a rough estimate of site traffic. We determine this via an algorithm that uses the Dataprovider.com map of the web and connects this to SEO data, Alexa and social media usage. If data from these sources is insufficient, this field will be left blank.		12%
	Analytics ID	104,769	Analytics ID refers to an account ID for Google Analytics. Our spider extracts the Google Analytics code from a website's homepage and presents it in our database. Google Analytics lets a website measure advertising ROI and tracks flash, video and social networking websites and applications.		11%
	AdSense ID	5,601	AdSense ID refers to the user ID of an AdSense account. Google AdSense is used to target certain audiences with advertisements in the form of text, image or video. Our spider uses this information to index how many websites a certain publisher has.		1%

섹션	변수	항목수	내용	정확도	범위
	Brand names	5,828	Our spider is trained to recognize popular brand names. These brand names often appear on the internet as counterfeit articles. The spider checks whether these brand names occur in the anchor texts of the hyperlink on the homepage. Brand names in combination with a low trust score mostly indicate a fraudulent website.		1%
	Analytics	102,968	Analytics refers to the type of analytics software that is used on a website to monitor the number of visitors and visitor behavior. Our spider checks the tracking code on a website and compares it to the codes from the Top 50 Analytics platforms that it knows for an accurate indexation.		11%
	Ad network	15,849	Ad Network refers to the network that is responsible for showing advertisements on a website. Our spider currently recognizes and detects over 20 Ad Networks by using the source code pattern of a website.		2%
	Affiliate	57	Affiliate refers to the company that tries to find publishers for advertisements. Our spider detects these affiliate companies based on the HTML code of the indexed website. Amazon Associates is one of the best know affiliates.		0%
	Social media types	100,041	Social media types shows the social media platforms that can be found on the indexed website. This field is not limited to the homepage, social media platforms on all indexed pages of the website are included in this field.	100%	11%
	Social media widgets	19,664	Social media widgets refers to the widget applications that are found on the homepage of a website. These widgets make it easier for visitors of the website to share a website page on their own social media profiles.		2%
	Social media profiles	100,041	Social media profiles are the addresses (URLs) of the social media profiles that are found on a website. This is not limited to the homepage, but contains all profiles found on the indexed pages of the website such as Facebook, Twitter, Instagram, Pinterest and several others.	87%	11%
	Live chat software	652	Live chat refers to the buttons or icons on a website that offer a live chat. These chats are often used for page visitors who have a question or a problem. Our spider detects live chat by analyzing whether a link to a known live chat platform is embedded on the homepage.		98%
	CRM	238	Customer relationship management (CRM) is the system a company uses to manage interaction with current and future customers. A CRM analyzes data about the customers' history with the company, in order to improve business relationships and drive sales.		0%
	Email services	453	Email services refers to the email management system a website uses. These services provide the hosting of servers that send, receive and store email for the website and allow the website to create personalized email addresses.		0%
	TLD suggestions	928,184	The TLD suggestions contain a list of TLDs that are a good fit for the domain. With machine learning and several data points taken from the website, we generate a short list of TLDs that are a good match for the domain. The TLD suggestions take into account the content, geographical location of the website, company name and subdomains.		100%
Technical			Where possible we give insight into the way a website is built and which tools it facilitates, for example the scripting language used and whether a website supports flash. This category also includes an evaluation of the technical aspects and the SEO score of a website.		
	CMS	55,319	Content Manager System (CMS) refers to the CMS that is used for the website. A CMS is an application that supports the creation and modification of digital content. Our spider currently detects more than 100 systems, by looking for specific CMS fingerprints in the HTML code.		6%
	Scripting language	627,174	Scripting language is the computer language that has been used to build the website. Our spider indexes the scripting language of a website by looking at its file extensions or the header, for example PHP or JavaScript.		68%

섹션	변수	항목수	내용	정확도	범위
	Technical evaluation	928,195	Technical evaluation is a scoring system we use to determine how well a website is built technically. We rate 34 technical factors on a scale of 1 to 10 for each website to determine a final score. Factors that are used are the use of flash or frame, the use of advanced techniques and several other variables.		100%
	SEO score	928,195	The SEO (Search Engine Optimization) score refers to how well the website is optimized for search engines. This score is measured in percentages and a higher percentage indicates a better optimized website. We determine this score by SEO guidelines that search engines such as Google and Bing have provided.		100%
	SEO summary	928,195	Our SEO summary presents the most important results of a complete SEO analysis of a website. This summary is shown in an encoded string, that can be decoded by an additional and optional Dataprovider.com script. Once you put the encoded string in the decoder, it will generate a text report on all major SEO topics.		100%
	Flash	928,195	Flash refers to the use of flash on a website. Flash is a software technology that is used for creating and managing interactive multimedia web applications, and is often used to show animations on the pages of a website.		100%
	HTML version	316,158	HTML version refers to the current HTML version that is being used for a website. Our spider determines the HTML version that is being used based on the DOC TYPE definition. The most current version is HTML 5.		34%
	Generator	87,760	Generator is a META tag. It contains information about the CMS that was used to build the website and the version of that CMS. Our spider detects this information by using the META description in the HTML code of the website.		9%
	Mobile version	928,195	Mobile version refers to the availability of a mobile optimized website. Our spider determines mobile optimization based on the information it receives when it tries to open a mobile website. If it receives the same content for both desktop and mobile, this indicates there is no mobile version available.		100%
	Mobile app	4,683	Mobile app refers to the use of a mobile store on the website. If a link to a mobile store is present, this is a likely indication that the website also offers a mobile application. We currently support iTunes App Store, Google Play Store and Windows Mobile Store.	95%	1%
	Maps	3,743	Maps refers to the presence of a map on the website. These maps are often used to indicate the location of a company's headquarters or to give more information about where the company is based. We currently index 8 different map applications.	97%	0%
	Libraries	233,261	The libraries field contains a list of all the default external libraries that can be used to add functionalities to the website. It is possible for a website to use multiple websites. Examples of libraries are jQuery, Twitter Bootstrap and Modernizr.		25%
	Scanrequest ID	928,218	The Scan Request ID is a unique code that Dataprovider.com attributes to each website we index. Our spider uses these codes internally to track an analysis and ensure that can provide information on monthly changes via Trends.	100%	0%
Hosting			Our hosting filters give extensive information on the hosting, domain and webserver of a website. Additionally, this category informs you on the presence of an SSL certificate, average load time of a website and whether a website is parked.		
	Hostname	928,218	The hostname is the complete address were a website is located. It consists of a subdomain, a domain and a top-level domain. For example: hostname maps.google.com includes domain 'google.com', a sub-domain 'maps' and a top-level domain 'com'. A domain can hold multiple sub-domains and thus multiple hostnames.		100%
	Top-level domain	928,218	A top-level domain (TLD) is the last segment of a domain name, the letter immediately following the final dot. In the domain name www.dataprovider.com for example, the top-level domain is .COM. Other examples or top-level domains are .NL and .INFO.		100%

섹션	변수	항목수	내용	정확도	범위
	Top-level domain type	928,218	Top-level domain type refers to the type of top-level domain that a website uses. This can be a Generic TLD (gTLD) such as .COM or .ORG, a Country Code TLD (ccTLD) such as .DE or .NL, a Sponsored TLD (sTLD) such as .MOBI or .COOP or a New Generic TLD (ngTLD) such as .ECO.		100%
	New Generic Top-level domain category	25,099	If a website uses a new top-level domain, this field will categorize the kind of new top-level domain that is being used. Categories are generic, geographic and brands.		3%
	Subdomain	880,734	In our database, subdomain refers to the first part of a domain. The most used sub domain in domain names is 'www' but there are other subdomains such as 'shop'. If no subdomain can be found, this field is left blank.		95%
	Domain	928,218	Domain refers to the body of the URL: the middle part of the entire domain name. This is a unique domain that services as an identifier for a website. For example: dataprovider.com is the domain for the website www.dataprovider.com.		100%
	Domain name length	928,218	Domain name length refers to the number of characters that is used in a domain name. The minimum length of a domain name is 1 character and the maximum length of a domain name is 255 characters. Both the subdomain and the top-level domain are not counted here.		100%
	IDN	928,218	An Internationalized Domain Name (IDN) is a domain name that contains at least one character that is displayed in a language-specific way, such as Arabic or Chinese. If the domain name of a website contains a language-specific character this field is true.		100%
	IDN parts	81,009	The Internationalized Domain Name (IDN) part refers to the part of the domain name that is using language-specific characters. These parts can be the subdomain, the middle part of the domain or the top-level domain.		9%
	Linked subdomains	20,407	Most common subdomains start with WWW, but it is possible to use an alternative subdomain such as blog.example.com. We call these linked subdomains. This field display the additional subdomains that were found during indexation.		2%
	Linked subdomains count	20,407	Linked Subdomain Count refers to the number of subdomains that was found on a website during the indexation by our spider. These linked subdomains are alternative subdomains such as blog.example.com or shop.example.com.		2%
	Forwarding domains	17,120	Forwarding domains are domains that redirect from one domain to another domain. Some domain name owner register variations of their domain and redirect these variations to the main domain name. Those variations are then called the forwarding domains.		2%
	Forwarding domains count	17,120	Forwarding domain count refers to the number of domains that redirects to the main domain via a server redirect. For example: hosteuropegroup.com, dastadock.de and intergenia.de all redirect to heg.com. This means heg.com has 3 forwarding domains.		2%
	Hosting country	928,195	Hosting country refers to the country where the website is hosted. Our spider determines the hosting country based on the IP Address of a website. The hosting country can differ from the country where the company behind the website is based.		100%
	IP address	928,218	IP Address refers to the number of the server that contains a website. Each website has an IP address. A website can have multiple IP addresses when it is hosted at multiple locations, this field can thus contain multiple addresses.		100%
	IPv6 address	11,791	IPv6 is the Internet's next-generation protocol that is designed to replace IPv4. This version allows more users and devices to communicate on the internet. If no IPv6 can be found this field will be left blank.		1%

섹션	변수	항목수	내용	정확도	범위
	AS number	928,152	The AS (Autonomous System) number corresponds with a company that is responsible for the network on which a website is hosted. Each website is hosted on an IP address and these IP addresses are stored in an AS. The AS number 8,560 for example means a website is hosted with 1 and 1 Internet SE.		100%
	Hosting company	928,152	The hosting company is also known as the AS company. This company is responsible for the network on which the website is hosted and owns the IP address the website is on. The hosting company corresponds with the assigned AS number.		100%
	Registrar	872,905	Registrar refers to a company that has the authority to register a domain name and top-level domain such as .COM, .INFO or .DE. These companies are accredited to sell and register a domain name and can offer additional services such as email services.		94%
	Reseller	188,957	Reseller refers to a company or person who manages and sells domain names that have been registered before, but is not an accredited registrar. Information on this reseller can also be found in our WHOIS data.		20%
	DNS NS domain	729,052	DNS NS (Name Server) Domain refers to the domain name of the primary server used in the DNS. The name server is a computer that is permanently connected to the internet and translates domain names into IP addresses. It contains the DNS of a domain with all IP addresses that belong to that domain.		79%
	DNS MX domain	317,048	The DNS MX domain is the mail exchanger record (MX). This MX domain specifies the mail server that is responsible for accepting and sending email messages for a domain, for example googlemail.com.		34%
	DNS TXT	226,205	A TXT record is a type of resource record in the DNS that can be used to assign free text to a hostname. The most common use for TXT records is to store Sender Policy Framework (SPF) records and to prevent fake emails being sent out via your account.		24%
	DNSSEC	735,847	The DNSSEC refers to a Domain Name System Security Extension. This extension is used for securing information that is provided by the DNS. Extensions are not mandatory and this field shows whether one is being used.		79%
	DNS NS Nameservers	734,170	DNS NS Nameservers refers to the nameservers (multiple) used in the DNS. The name server is a computer that is permanently connected to the internet and translates domain names into IP addresses. It contains the DNS of a domain with all IP addresses that belong to that domain.		79%
	Operating system	299,109	Operating system refers to the type of operating system being used to generate the website. Our spider looks for an operating system in the header of a website. Not every server returns operating system information because it can pose a security risk, if that is the case this field will be blank.		32%
	Webserver	610,763	Webserver refers to the server that is used to store, process and deliver the pages of a website to its clients. Each website needs a webserver, but not every server returns this information because it can pose a security risk. If the information is not returned, this field will be blank.		66%
	Control panel	1,391	A control panel allows a server administrator to set up new websites, reseller accounts, email accounts, and DNS entries through a web-based interface. The administrator can create client and site templates, which predetermine resource-allocation parameters for the domains and/or clients. Plesk and cPanel are the most popular control panels.		0%
	Server signature	896,886	The server signature contains information about the operating system and the web server that a website uses. Not every server returns this information because it can pose a security risk, if that is the case this field will be left blank.		97%

섹션	변수	항목수	내용	정확도	범위
	SSL certificate	928,213	SSL certificate refers to whether the padlock and the 'https' domain can be seen in a browser. This certificate is used to validate the website's security of message transmission on the web and is important with for example credit card transactions and online logins.		100%
	SSL issuer organization	67,304	The certificate issuer organization is responsible for delivering the certificate to a website. An SSL certificate is a digital certificate that authenticates the identity of a website and encrypts information sent. These certificates can only be issued by a Certificate Authority.		7%
	SSL issuer common name	67,304	The SSL issuer common name refers to the name of the certificate issued by the SSL issuer organization. An SSL certificate is a digital certificate that authenticates the identity of a website and encrypts information sent.		7%
	SSL start date	67,304	An SSL certificate is a digital certificate that authenticates the identity of a website and encrypts information sent. These certificates are only valid for a definite period. If an SSL certificate is found, this field shows the start date of that certificate.		7%
	SSL end date	67,304	An SSL certificate is a digital certificate that authenticates the identity of a website and encrypts information sent. These certificates are only valid for a definite period. If an SSL certificate is found, this field shows the end date of that certificate.		7%
	SSL RSA key length	67,304	The key length refers to the length of an RSA public key. This length gives an indication of the strength of the encryption that is used for an SSL certificate. A key length of 2048 or 4096 is currently seen as secure.		7%
	SSL algorithm	4	SHA stands for 'Secure Hash Algorithm' and is used by certification authorities to sign certificates. There are several versions of SHA including SHA-224, SHA-256, SHA-384,SHA-512,SHA-512/224 and SHA-512/256. Websites using SHA1 are more susceptible to attacks.		0%
	SSL type	67,304	There are three types of SSL certificate available: domain validation (default), organization validation or extended validation. Currently extended validation is seen as the most secure type of SSL certificate, it can be recognized by a green address bar in your browser.		7%
	Status codes	928,210	A status code is the HTTP code used on a website when the request cannot be fulfilled. Our spider shows the status codes that are available on a website. The best-known status code is 404 for PAGE NOT FOUND.		100%
	Average load time (ms)	926,229	Average load time refers to the time it takes for a webpage to load. Our spider measures this from the time it asks for a page until the page is received. Average load time is shown in milliseconds (MS).		100%
	CDN	38,536	Content Delivery Network (CDN) refers to the distribution system on the web where a website stores its files. If you download a whitepaper from a website for example, it is very likely that this file has been stored in the CDN.		4%
	Video	22,787	Video refers to the use of external sources on a website to show videos. Our spider can currently index over 20 types of video players. The most used video players are YouTube, Vimeo and JM Player.	99%	2%
	Parking	42,591	Parking refers to a registered domain name that does not contain a website. A parked domain may have been registered for future development or to protect the domain from the possibility of cybersquatting. Our spider determines parked pages by looking at the HTML code and Name Server.	88%	5%
	Placeholder	126,851	A placeholder is a temporary default homepage generated by the hosting company of a website. Such a placeholder is usually an indication that a permanent website is being created. Our spider detects default placeholder pages from all major registrars, as well as customized placeholders that use default text.	81%	14%

자료: 데이터프로바이더(www.dataprovider.com)