

논문 2014-51-7-17

특허분석을 통한 빅 데이터의 시각화 기술 분석

(Analysis of Big Data Visualization Technology Based on Patent Analysis)

노 승 민*, 최 용 수**

(Seungmin Rho and YongSoo Choi)

요 약

현대 데이터 컴퓨팅의 발전은 그래픽 기능의 향상을 이끌고 데이터 디스플레이를 위한 많은 가능성을 가지고 있다. 시각화는 방대한 양의 데이터에서 중요한 정보를 나타내줄 뿐만 아니라 복잡한 분석 방식에 효율적임이 증명되었다. 빅 데이터 분석과 발견은 컴퓨터 그래픽과 시각화 커뮤니티에서 새로운 연구 기회로 제시되고 있다. 본 논문에서는 주요 시장 국가의 특허 분석을 통해 빅 데이터의 시각화 기술 개발 동향을 살펴보고자 한다. 특히, 2012년 11월을 기준으로 4개국에 출원 및 등록되어 있는 160건을 대상으로 분석을 진행하였다. 분석결과에 따르면 텍스트 클러스터링, 2D 시각화 분야의 기술개발이 중요하며 이에 대한 시급한 개발을 지향할 필요가 있다. 특히, 국내에서 다양한 스마트 디바이스의 증가와 소셜 네트워크 사용 증대로 인해 빅 데이터 입체 시각화 기술 개발이 매우 시급함을 알 수 있다.

Abstract

Modern data computing developments have led to big improvements in graphic capabilities and there are many new possibilities for data displays. The visualization has proven effective for not only presenting essential information in vast amounts of data but also driving complex analyses. Big-data analytics and discovery present new research opportunities to the computer graphics and visualization community. In this paper, we discuss the patent analysis of big data visualization technology development in major countries. Especially, we analyzed 160 patent applications and registered patents in four countries on November 2012. According to the result of analysis provided by this paper, the text clustering analysis and 2D visualization are important and urgent development is needed to be oriented. In particular, due to the increase of use of smart devices and social networks in domestic, the development of three-dimensional visualization for Big Data can be seen very urgent.

Keywords : Big data, Patent, Visualization, Graphical Representation, Three-Dimensional

I. 서 론

세계 경제 포럼은 2012년 떠오르는 10대 기술 중 그 첫 번째를 빅 데이터 기술로 선정했으며 대한민국 지식경제부 R&D 전략기획단은 IT 10대 핵심기술 가운데 하나로 빅 데이터를 선정하는 등 최근 세계는 빅 데이터를 주목하고 있다^[5].

디지털 경제의 확산으로 우리 주변에는 규모를 가늠

* 정회원, 성결대학교 멀티미디어공학과
(Dept. of Multimedia Eng., Sungkyul University)

** 정회원, 성결대학교 교양교직부
(Div. of Liberal Arts & Teaching Sungkyul University)

© Corresponding Author(E-mail: ciechoi@sungkyul.ac.kr)

접수일자: 2014년04월30일, 수정일자: 2014년05월22일
수정완료: 2014년06월24일

할 수 없을 정도로 많은 정보와 데이터가 생산되는 '빅 데이터(Big Data)' 환경이 도래하고 있다. 빅 데이터란 과거 아날로그 환경에서 생성되던 데이터에 비하면 그 규모가 방대하고, 생성 주기도 짧고, 형태도 수치 데이터뿐 아니라 문자와 영상 데이터를 포함하는 대규모 데이터를 말한다. 위키 백과에서 정의하는 빅 데이터는 "기존 데이터베이스 관리도구로 데이터를 수집, 저장, 관리, 분석할 수 있는 역량을 넘어서는 대량의 정형 또는 비정형 데이터 집합 및 이러한 데이터로부터 가치를 추출하고 결과를 분석하는 기술"이다^[6]. 빅 데이터로의 발전은 정보사회의 패러다임을 견인할 정도의 큰 힘을 발휘하고 있음은 누구도 부인하기 어렵다. 스마트 단말기 확산, SNS 활성화, 클라우드, M2M(Machine to Machine) 네트워크의 확산으로 데이터 폭발이 더욱 가속화되어 빅 데이터의 기반이 확대 되고 있다. 따라서 앞으로 5년 이내 빅 데이터를 위한 고급분석 등 관련 기술은 성숙될 것으로 전망하고 있다^[1].

다양한 종류의 대규모 데이터의 생성·수집·분석·표현을 그 특징으로 하는 빅 데이터 기술의 발전은, 다변화된 현대 사회를 더욱 정확하게 예측하여 효율적으로 작동케 하고, 개인화된 현대 사회 구성원마다 맞춤형 정보를 제공·관리·분석 가능케 하며, 과거에는 불가능했던 기술을 실현시키는 고부가가치 기술로 특히 빅 데이터의 입체 시각화 기술은 대량의 다양한 데이터 생산 기술의 진보와 이에 필요한 데이터 저장·관리·분석 기술의 발전 속에서 필수적으로 요구되는 기술이다^{[2][3]}.

본 논문에서는 빅 데이터의 입체 시각화 기술을 개발함에 있어 비정형 데이터로 한정하지 않고 텍스트 분류, 텍스트 클러스터링, 네임드 엔티티(Named Entities), 정형/비정형 텍스트의 평판분석, 의미 추출 기술 등을 포함하는 텍스트 마이닝 기술과, 빅 데이터를 2D, 3D로 시각화하는 기술 그리고 빅 데이터를 심미적인 틀이 추가된 시각화기술을 포함하는 시각화 기술에 대하여 특허동향분석을 통하여 국제 특허현황 및 국가별 기술 경쟁력 등의 분석하여 최근 부상기술 등을 도출하였다.

II. 특허 분석 범위

본 논문에서는 연구 성과의 파급효과 및 연구의 필요성을 고려하여 선택된 기술 분야를 특허분석대상으로 하였으며, WIPSON DB 기반의 1991년 1월~2012년 11

월까지 출원공개 된 한국, 일본, 유럽 및 미국 공개특허와 1991년 1월~2012년 11월까지 출원등록 된 미국등록 특허를 분석 대상으로 하였다.

2.1. 특허 분석대상 기술

분석대상 기술은 빅 데이터의 시각화 기술을 위해 텍스트 마이닝, 시각화 기술 두 가지 범주로 나누었고, 세부 기술로는 텍스트 분류, 클러스터링, 정보추출, 정형/비정형 텍스트의 평판분석, 의미추출, 빅 데이터 시각화(2D, 3D, 심미적 시각화)로 분석하였다. 표 1은 특허 분석대상의 기술 분류와 범위를 나타낸 것이다.

표 1. 분석대상 기술 분류
Table 1. Analyzed the technology classification.

중분류	소분류	검색개요 (기술범위)
텍스트 마이닝	텍스트 분류	<ul style="list-style-type: none"> ● 비정형 데이터에 한정하지 않음 ● 데이터 혹은 빅 데이터의 텍스트를 분류하는 기술
	클러스터링	<ul style="list-style-type: none"> ● 비정형 데이터에 한정하지 않음 ● 데이터 혹은 빅 데이터의 분류된 텍스트를 정해진 범주별로 분류하는 기술
	정보추출	<ul style="list-style-type: none"> ● 비정형 데이터에 한정하지 않음 ● 데이터 혹은 빅 데이터의 텍스트가 가진 개념을 추출해내는 기술
	평판분석	<ul style="list-style-type: none"> ● 비정형 데이터에 한정하지 않음 ● 정형(구조적)/비정형(비구조적)텍스트를 분석함으로써, 텍스트가 의하는 의견(의도) 혹은 정서를 분석하는 기술을 포함함
	의미추출	<ul style="list-style-type: none"> ● 비정형 데이터에 한정하지 않음 ● 데이터 혹은 빅 데이터의 분류, 클러스터링을 통해 추출된 단어의 의미를 추출해내는 기술 ● ex. 레스토랑과 식당이 같은 의미임을 추출함
시각화	2D	<ul style="list-style-type: none"> ● 텍스트 마이닝을 실시한 빅 데이터를 평면적으로 시각화하는 기술
	3D	<ul style="list-style-type: none"> ● 텍스트 마이닝을 실시한 빅 데이터를 3D 혹은 입체적으로 시각화하는 기술
	심미적 시각화	<ul style="list-style-type: none"> ● 텍스트 마이닝을 실시한 빅데이터를 시각화함에 있어서, 심미적인 효과에 대한 내용이 기술

2.2. 특허 검색결과

분석 대상 기술에서 유효특허를 추출해내기 위해 각 기술에서 노이즈를 제거한 후 선별된 결과는 표 2와 같고, 유효데이터는 한국, 미국, 일본, 유럽에 특허 출원된 것을 기준으로 추출하였다.

표 2. 유효특허 추출결과
Table 2. Effective patent extraction results.

중분류	소분류	유효데이터 건수				
		한국	미국	일본	유럽	계
텍스트 마이닝	텍스트 분류	6	12	4	2	24
	텍스트 클러스터링	5	13	25	5	48
	정보추출	1	12	26	2	41
	텍스트의 평판분석	8	6	5	0	19
	의미추출	2	1	2	0	5
소 계		22	44	62	9	137
시각화	시각화(2D)	0	7	7	1	15
	시각화(3D)	0	7	1		8
	심미적 시각화	0	0	0	0	0
소 계		0	14	8	1	23
총 계		22	58	70	10	160

III. 특허기술의 Landscape

3.1. 국가별 Landscape

빅 데이터의 입체 시각화 기술 개발 분야의 연도별 전체 특허동향을 살펴보면, 거시적인 관점에서 특허출원건수는 1992년부터 2003년까지 완만한 증가세를 보이다가, 2004년~2005년에 일시적으로 출원건수가 감소하였다가 이후 다시 증가 추세를 보이고 있다. 미공개 특허가 존재하는 2011년 및 2012년 구간도 소셜 미디어의 성장과 최근 스마트폰으로 대변되는 모바일 장치의 확산이 결합됨에 따라 일상생활에서 대규모 데이터가 급증할 것이며, 이에 따른 빅 데이터 분석 분야에 대한 관심이 높아질 것을 고려해볼 때, 증가추세 곡선이 예측된다.

한국은 2000년도에 최초출원을 시작으로 2006년까지는 비정형 빅 데이터의 입체 시각화 기술 개발 분야에 출원활동이 미비하였으나, 2007년 이후 지속적인 증가

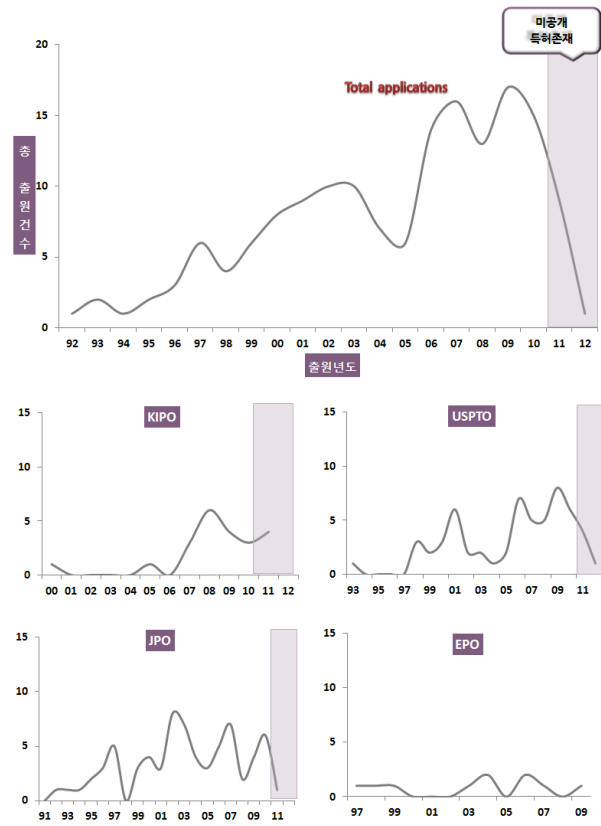


그림 1. 주요 시장 국 연도별 특허동향
Fig. 1. Patents of major market countries by year.

추세를 보인다. 이는 한국은 데이터 기반의 사업 환경이 미성숙한 상황이기 때문이며, 타국에 비해 연구개발의 시점은 다소 늦으나, 추후 활발한 연구 활동이 있을 것으로 분석된다.

미국은 전체 출원동향과 유사한 출원동향을 보이고 있으며, 2003년~2005년 사이 출원건수가 일시적으로 감소하였으나, 2006년 이후 지속적인 출원활동을 하고 있는 것으로 조사된다.

일본은 분석구간 초기에는 활발한 출원활동을 지속하다가, 최근 소폭 감소세를 나타내는 것으로 나타났으며, 유럽은 소수의 특허를 지속적으로 출원하고 있는 것으로 분석된다.

빅 데이터의 입체 시각화 기술 개발 분야의 관련특허는 2000년대 후반에 꾸준한 증가추세를 보이고 있으며, 이는 스마트폰 사용자의 증가와 소셜 네트워크 서비스의 사용증대로 인한 빅 데이터의 증가하고, 빅 데이터 분석에 관한 필요성이 증가함에 따른 결과인 것으로 판단된다.

3.2. 주요 시장 국가의 특허출원 현황

빅 데이터의 입체 시각화 기술 개발 분야의 국가별/출원인 국적별 특허동향을 살펴보면, 일본에서의 출원이 전체 분석대상 국가 출원점유율의 약 44%로 가장 많은 특허를 출원하였으며, 다음으로 미국에서의 출원이 전체 출원점유율의 약 36%인 것으로 나타나, 비정형 빅 데이터의 입체 시각화 기술 개발 분야의 연구개발은 대부분 일본과 미국에서 주도되고 있는 것으로 보인다.

주요시장국의 내·외국인 특허출원현황을 살펴보면, 일본과 미국에서의 외국인 점유율은 20%, 17% 로 외국인 보다는 내국인에 의한 특허출원이 활발한 것으로 나타났고, 한국에서의 외국인 출원점유율은 9%로 타 국가에 비해 외국인의 특허활동이 매우 저조한 것으로 조사되었으며, 그 중 미국 국적의 출원인만이 한국에 출원활동을 하고 있는 것으로 분석되었다. 이와 같은 추세를 판단할 때, 한국 국적의 출원인들이 내국 위주의 특허활동을 하고 있는 동시에, 타 주요시장국에서 한국 시장에 진출하는 비율이 높지 않기 때문인 것으로 판단되며, 이를 볼 때, 해외 국적의 출원인들이 한국의 시장성을 미국 및 유럽에 비하여 상대적으로 더 낮은 것으로 판단하여 한국 시장에 적극적으로 진입하지 않고 있기 때문인 것으로 볼 수 있다.

유럽에서의 외국인 점유율이 80%로 외국인에 대한 의존도가 다소 높은 것으로 나타났으며, 이중 미국 출

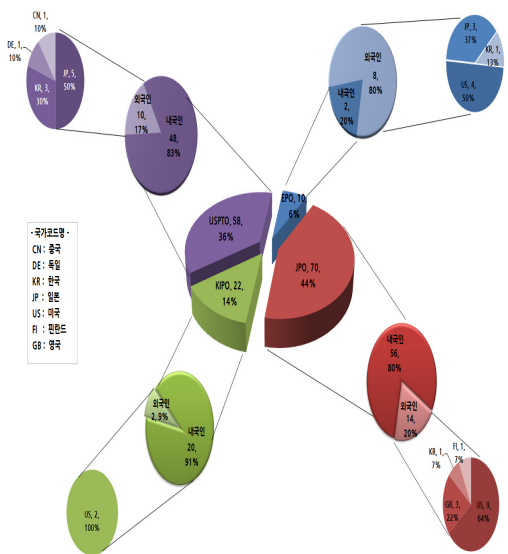


그림 2. 주요 시장 국 특허출원동향
Fig. 2. Patents application of major market countries.

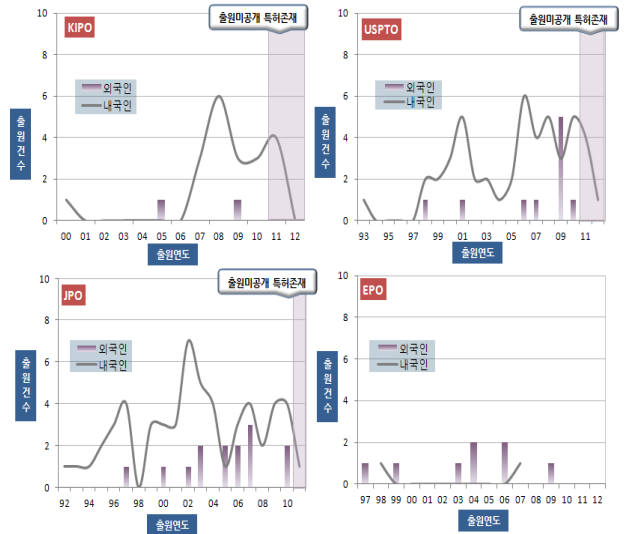


그림 3. 연도별 주요 시장 국 특허출원동향
Fig. 3. Patents application of major market countries by year.

원인이 50%, 일본 출원인이 37%, 한국출원인이 13%의 출원점유율을 나타내며, 유럽에서 출원활동을 하고 있는 것으로 보인다.

미국에서의 내·외국인 연도별 출원동향을 살펴보면, 출원 초반에 내국인의 특허출원이 증가세를 보이다가, 2002~2005년 사이에 출원건수가 감소하였으나, 다시 증가추세를 보이고 있지만, 외국인의 출원은 2009년에 일시적으로 급증한 것 이외에는 소수의 특허를 출원하고 있는 것으로 판단할 때, 미국특허는 내국인 위주로 연구개발 활동이 이루어지고 있는 것으로 분석할 수 있다.

특허출원 점유율이 2위인 일본의 외국인 연도별 출원동향을 살펴보면, 미국 국적의 출원인이 64%로 높은 점유율을 나타내고 있으며, 다음으로 영국, 한국, 핀란드 국적의 출원인이 일본에 특허출원을 하고 있는 것으로 나타나고, 또한 최근 외국인의 출원건수가 소폭 증가하고 있는 것으로 보아 일본시장에서의 외국출원인의 활동에 대한 지속적인 모니터링이 필요할 것으로 판단 된다.

3.3. 시장별 세부기술 동향

각국의 특허청에 출원된 출원 데이터를 기준으로 세부기술의 집중도 및 공백영역 등을 버블그래프로 나타내어 해당 시장의 관심도를 나타내고자 한다. 관심 있는 시장에서 어떠한 세부기술이 관심 있게 특허출원 되

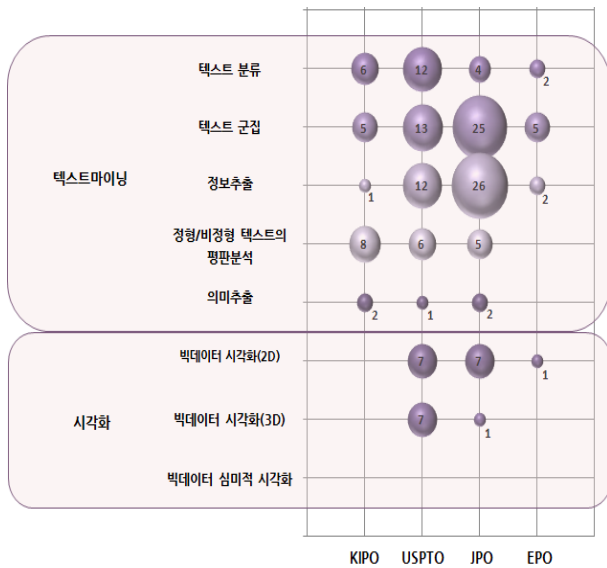


그림 4. 시장별 세부기술동향
Fig. 4. Details technology trends by market.

고 있는가를 파악하고자 한다. 또한 해당 세부기술에 대한 시장별(특허청별) 비교 분석도 가능하다.

한국은 빅 데이터의 입체 시각화 기술 개발 분야에서 텍스트 마이닝 분야에는 특허출원건수를 보유하고 있으며, 텍스트 마이닝 분야에서 정형/비정형 텍스트의 평판분석은 다른 기술 분류와 비교에 많은 특허출원건수를 보유하고 있다. 시각화 분야에는 연구 활동이 아직 없으며, 이는 한국이 타국과 비교하면 최초 출원이 늦으며, 아직 시각화 분야에 대한 관심이 시작된 지 얼마 안 되었기 때문으로 분석된다.

일본은 세부적으로 텍스트 분류 분야, 텍스트 군집 분야, 정보추출 분야에 각각 12건으로 다수의 특허를 출원하고 있고, 빅 데이터 시각화(2D) 분야와 빅 데이터 시각화(3D) 분야에도 활발한 연구 활동을 진행 중에 있어, 시각화 분야에는 다른 타 국가보다 일본의 연구가 활발한 것으로 판단된다.

미국은 특허출원건수가 70건으로 비교적 높은 수치를 나타내고 있는데, 이는 주요출원인인 Microsoft와 IBM가 텍스트 군집 분야와 정보추출분야에 대해 기술 개발과 다량의 출원으로 해당 기술 분야의 시장을 리드하고 있는 것으로 판단되며, 이러한 세부 기술별 특허출원 집중도는 일본에서의 추세와 유사한 것으로 분석된다. 유럽은 특허출원건수가 10건으로 매우 작은 수치를 차지하고 있다.

전체적인 각 시장국가별 세부기술 동향은 한국은 정

형/비정형 텍스트의 평판분석 분야, 미국은 텍스트 군집 분야와 정보추출 분야, 일본과 유럽은 텍스트 군집 분야에 집중적으로 출원한 것으로 판단된다.

IV. 결 론

빅 데이터라는 말이 유행처럼 번지는 시대 데이터 커뮤니케이션의 중요성이 날로 커지고 있다. 수많은 정보들이 범람하는 시대에 모든 정보가 의미 있는 것은 아니다. 원하는 정보를 줄 수 있어야 진정한 의미를 갖고 쉽게 전달되어야 가치가 있는 것이다. 최근 빅 데이터와 모바일의 트렌드에 맞추어 데이터의 시각화의 중요성이 점점 더 커지고 있다. 이러한 흐름에 맞추어 본 논문에서는 빅 데이터의 시각화 기술개발의 특허동향을 분석하기 위해 1991년 1월~2012년 11월까지 공개된 한국, 일본, 유럽, 미국 공개특허와 1991년 1월~2012년 11월까지 등록된 미국특허를 분석하였다.

분석결과 텍스트 마이닝 분야에서는 텍스트 클러스터링의 출원빈도가 높고, 시각화에서는 2D 시각화 분야의 출원 빈도가 제일 높음을 알 수 있었다.

스마트폰 사용자의 증가와 소셜 네트워크 서비스의 사용증대로 인한 빅-데이터 입체 시각화 기술 개발 분야의 관련특허는 2000년대 후반에 꾸준한 증가추세를 보이고 있는 것으로 분석되었고, 2D에만 국한되었던 시각화기술이 점차적으로 3D 시각화로 발전하고 있어 이에 대한 기술 및 연구개발이 이루어져야 할 것이다.

REFERENCES

- [1] Mckinsey Global Institute, "Big Data: The next frontier for innovation, competition, and productivity", 2011.
- [2] Julie Steele and Noah Illinsky, "Beautiful Visualization-Looking at Data through the Eyes of Experts," O'Reilly Media, 2012
- [3] "Free Data Visualization Software," tableau public, <http://www.tableausoftware.com/public/>
- [4] John Gantz & David Reinsel, "Extracting Value from Chaos," IDC IVIEW, June 2011.
- [5] Ministry of Knowledge Economy, "Hurried Steps for Leading IT Industry in Convergence.Smart Era ," MKE Press Release, Aril 2012.
- [6] <http://ko.m.wikipedia.org/wiki/>

— 저 자 소 개 —



노 승 민(정회원)
 2008년 아주대학교 정보통신
 공학과 (공학박사)
 2008년~2009년 Carnegie Mellon
 University, 박사 후 연구원
 2009년~2011년 고려대학교 전기
 전자전파공학부, 연구교수
 2012년~2013년 백석대학교 정보통신공학과,
 조교수
 2013년~현재 성결대학교 멀티미디어 공학과,
 조교수
 <주관심분야 : 음악 검색 및 추천, 집단 및 군집
 지성, 시맨틱 웹, 빅데이터>



최 용 수(정회원)-교신저자
 1998년 강원대학교
 제어계측공학과 공학사
 2000년 강원대학교
 제어계측공학과 공학석사
 2006년 강원대학교
 제어계측공학과 공학박사
 2006년~2007년 연세대학교 첨단융합건설연구단
 연구교수
 2007년~2013년 고려대학교 정보보호대학원
 연구교수
 2013년~현재 성결대학교 교양교직부(멀티미디어)
 조교수
 <주관심분야 : Multimedia Hashing, Information
 Hiding, Watermarking, Steganography, Image
 Forensics, Forgery Detection>