

## IPA를 이용한 개인정보 위험도 분석 연구

### A Study on Analysis of Personal Information Risk Using Importance-Performance Analysis

정수진\*, 김인석\*\*

Su-Jin Jeong\*, In-Seok Kim\*\*

**요약** 정보통신계의 발달로 인하여 등장하는 새로운 기술들로 인하여 새로운 개인정보의 형태가 나타나고 있다. 이에 따라, 기존 개인정보들과 결합되어 사용되어지는 개인정보들이 점차 늘어나고 있다. 현행 개인정보보호법에서 정의하는 결합된 정보에 대한 개인정보 위험도를 측정하는 방법은 정성적으로 제시되고 있어, 개인정보 위험도가 평균화되기는 어렵다. 본 논문에서는 기존 연구된 개인정보 위험도 평가 방법을 기반으로 개인정보 중요도와 가중치를 측정한다. 다음 IPA를 통해 개인정보의 위험도를 분석하는 모델을 제시하는데 그 목적을 두고 있다. 본 연구를 통하여 사용자의 주관적인 판단을 배제할 수 있고, 결합된 개인정보 위험도 산정에 사용될 수 있다. 또한, 제시되는 정략적인 위험도는 객관적인 지표로 사용될 수 있는 기준을 제시할 수 있을 것이다.

**Abstract** For newly incoming technologies owing to the advancements in information and communications technology, the new form of information occurs due to the combination form of personal information. In turn, personal information which is combination with existing personal information is on the increase. It is difficult to equalize the method in order to analysis the degree of risk for personal information because it is qualitative method which is defined on the current Personal Information Protection Law. This dissertation presents the model to assess the degree of risk by using the IPA(Importance-Performance Analysis) after measuring the importance and the weighted value for the personal information based on the existing the method of risk assessment. Through the model suggested in this dissertation, the subjective judgement can be excluded, the combination of personal information can be assessed and the standard criteria which is used as the objective indicators from the quantitative degree of risk can be suggested.

**Key Words** : Personal Information, risk, Security, TF-IDF, IPA(Importance-Performance Analysis)

## 1. 서론

정보통신계의 발달에 따라 새로운 정보기술들이 등장하고 있으며, 이러한 정보 기술들 간의 컨버전스 현상이 나타나고 있다. 정보처리 기술의 발전으로 사회 모든 분야에서 데이터 처리를 통한 부가가치 창출 사례는 증가하였으나, 데이터 처리 과정에서 개인정보의 결합을 통

한 재식별화, 민감정보 노출 등 개인정보 침해 가능성도 증가하였다.<sup>[1]</sup>

기술에 발전에 따라 그 가치가 중요시되고 있는 개인 정보란 살아 있는 개인에 관한 정보로서, 개인의 일반적 인 정보뿐만 아니라, 부동산정보·소득정보·고용정보·습관 및 취미정보 등 개인에 관한 사실 판단 평가를 포함한다. 또한, GPS 및 소셜정보를 통하여 개인의 위치 파악

\*준회원, 고려대학교 금융보안학과

\*\*정회원, 고려대학교 정보보호학과(교신저자)

접수일자: 2015년 10월 12일, 수정완료: 2015년 11월 12일

게재확정일자: 2015년 12월 11일

Received: 12 October, 2015 / Revised: 12 November, 2015 /

Accepted: 11 December, 2015

\*\*Corresponding Author: iskim11@korea.ac.kr

Dept. of Information Security, Korea University, Korea

을 할 수 있을 뿐만 아니라, 더불어 소비 형태와 금융 거래 정보까지 알 수 있는 새로운 형태의 정보도 포함한다.<sup>[2]</sup> 이러한 개인 정보들이 결합하면서 사회 변화에 따라 다양한 개인정보들이 생겨나고 있다.<sup>[3]</sup>

이처럼 개인정보의 범위가 확대 되면서 바이오 정보 뿐만 아니라 디지털 정보, 개인의 취향이나 사상까지 널리 퍼지고 있다. 이러한 개인의 사상을 가지고 개인을 구별해 낼 수 있고, 개인 침해 사고 사례가 점점 늘고 있다. 이러한 위협이 될 수 있는 많은 개인정보가 정부 시스템 뿐만 아니라 크고 작은 기업의 정보시스템에 의하여 쉽게 수집 및 축적되어 활용되고 있다. 이러한 개인정보가 유출되어 사고로 이어지고 있는데, 최근 발생하고 있는 개인정보유출 사고는 더 이상 개인의 피해만으로 끝나지 않으며 기업의 가치와 신뢰도에 큰 영향을 미치게 되었다.<sup>[4]</sup>

개인정보를 다루고 있는 모든 영역에서 개인정보 위험 관리에 대한 필요성이 대두되면서 한국인터넷진흥원에서 정보보호관리체계 위험관리 가이드라인을 고시하였다.<sup>[5]</sup> 그러나, 개인정보의 중요 정도가 얼마인지에 대해 정량적으로 측정이 미비하여, 각각의 개인정보의 위험도에 대한 정량적인 정보를 알고 있지 못하고 있는 실정이다.

본 논문에서 개인정보의 중요도와 TF-IDF를 이용하여 계산된 가중치를 토대로 개인정보 위험도를 분석하고자 IPA 기법을 응용하고자 한다. 본 논문 제2장에서는 관련연구에 대해 알아보고 제 3장에서는 기존의 개인정보 위험도에 대한 관련연구를 토대로 TF-IDF와 IPA에 대해 연구방안을 모색한다.

제 4장 5장에서는 기존 제시되어 있는 개인정보 결합도에 따른 중요도를 기반으로 각 개인정보별 중요도를 측정한다. 그 다음 각 금융사에서 수집하는 개인정보를 파악하여 TF-IDF 가중치 값을 구한다. 측정된 개인정보 중요도와 가중치를 가지고 IPA의 사분면을 이용하여 위험도를 정량적으로 측정하는 방안을 제시한다. 마지막으로 제 6장에서는 연구 결과에 대한 평가와 결론 및 향후 연구방향에 대해 정리하고자 한다.

## II. 관련 연구

### 1. 개인정보 위험도

개인정보 위험도란 외부의 위협이 개인정보 자산 내

부에 존재하는 취약성을 이용하여 자산을 훼손시키는 행위를 말한다. 잠재 위협요인이 위협의 발생 빈도와 개인 정보 자산의 손실크기를 조합하여 위험도를 결정하고, 더 나아가서 결정된 위험도를 통하여, 지속적으로 유해하는 위협을 찾아내거나 개선하는 종합적이고 체계적인 방법을 말한다.

또한 개인정보 라이프 사이클인 수집·저장·이용·제공·파기에서 개인정보와 관련된 사업을 추진 및 수행할 때, 고객의 개인정보 보호를 위하여 유출 및 오남용으로 인한 개인정보 침해가 없는지를 조사·예측·검토하여 침해요인을 제거하거나 최소화하는 절차를 개인정보 위험관리라고 한다.<sup>[6]</sup>

개인정보도 자산으로 보는 기존 연구에서는 개인정보 유출 시에 위험도를 평가하는 방법이 연구가 많이 되었다.

### 2. 개인정보 위험도 측정

한국인터넷 진흥원의 공공기관 개인정보 영향평가 수행 안내서에서 위험도는 개인정보 처리업무의 중요도, 침해요인의 발생가능성, 법률 규정된 의무사항 등 자산의 가치와 침해요인을 종합적으로 고려하여 도출한다. 또한 법률에 규정된 의무사항일수록 개인정보 침해 위험도가 크다는 것을 고려한 법적 준거성을 포함하여 위험도 산정하는 방안으로써, 다음 식(1)과 같다.<sup>[7]</sup>

$$\begin{aligned} \text{위험도} = & \text{자산가치} + (\text{위협 측면의 침해요인} * \text{법적 준거성}) \\ = & \text{자산가치} (\text{개인정보영향도}) \\ & + (\text{침해요인 발생 가능성} * \text{법적 준거성}) * 2 \end{aligned} \quad (1)$$

개인정보보호를 위한 개인정보 유출 모니터링 시스템의 설계(조성규, 전문석)에서는 개인정보 유출과 관련된 요소들을 핵심위험 지표들로 수치화하여 관리 하는 개인정보 유출 모니터링 시스템의 설계 방안에 대해 제시하였다.<sup>[8]</sup>

개인정보 노출을 방지하기 위한 위험도 모델에 대한 연구(이기성, 안효범, 이수연)논문에서는 각 개인정보의 노출 정도(EL)와 자산가치(P)를 곱하여 위험도(RV)를 추출해 내는  $RV = EL * P$ 를 제안했다.<sup>[9]</sup> 또한, 빅데이터 개인정보 위험 분석 기술(최대선, 김석현, 조진만, 진승현) 논문에서는 빅데이터 분석에 따른 프라이버시 침해 가능성과 이에 대응하기 위해 빅데이터상의 노출된 개인정보 위험도를 분석하는 기술을 제안 하였다.<sup>[3]</sup>

SNS에서 개인 정보 유출 방지를 위한 개인정보 유출 위험도 측정 방법(천명호, 최종석, 신용태) 논문에서는 개인정보 자산가치와 관계를 기반으로 SNS에서 각 개인정보별 노출 빈도율과 접근율을 통하여 개인정보 유출 위험도 측정방법에 대해 제안 하였다.<sup>[10]</sup> 개인정보를 세분화 후, 각 개별 위험도를 산정하고 개인정보가 복합될 경우의 피해와 위험을 고려하여 개인정보별 노출시 피해를 정량적으로 산출할 수 있는 위험도를 개인정보 노출에 대한 정량적 위험도 분석 방안(김평, 이윤호, 티무르쿠다 이베르케노프) 논문에서 정의하였다.<sup>[11]</sup>

기존 연구들을 토대로 보면 Privacy 측면에서 개인정보의 가치 면에서 보면 허가 받지 않은 자가 정보를 봤을 때 위험이 발생한다. 즉 기밀성이 결여되거나 손상을 입었을 때 피해와 영향이 발생하므로 개인정보는 기밀성에 의존하는 특성을 가진다. 이처럼 개인정보 위험도를 측정을 위한 연구가 다수 되었으나, 기존 연구에서는 실제 위험도를 판단하는 기준이 정량적 제시된 연구가 미비하여, 본 논문에서는 개인정보 위험도를 정량적 제시하는 방안을 제시 한다.

### III. 개인정보 위험도 측정 방법

#### 1. TF-IDF 가중치 모델

가중치 계산 기법 중 어휘 빈도수(Term Frequency) 계산법을 응용하여, 개인정보 수집 시 개인정보 항목이 수집된 빈도를 계산한다. TF-IDF 가중치 모델이란 1972년에 Sparck Jones에 의해 제안되었으며, 원래 목적은 문헌빈도(Inverse Document Frequency)에 용어 빈도를 곱하여 단어에 가중치를 부여하는 용도로 사용되었다.<sup>[12]</sup>

TF-IDF 모델에서 TF는 단어의 출현빈도가 문서 내에 대한 단어의 중요성을 나타내는 값으로, 단어 가중치 값 부여에 있어 가장 단순한 방법이다. TF값이 높을수록 해당 단어가 문서에서 중요하다고 정의한다. 일반적으로 특정 문서  $d_j$ 에서 단어  $t_i$ 의 중요도는 다음 식(2) 같이 표준화된다.<sup>[13,14]</sup>

$$tf_{i,j} = \frac{n_{i,j}}{\sum_k n_{k,j}} \quad (2)$$

$n_{i,j}$  : 문서  $d_j$ 에서 특정 단어  $t_i$ 의 출현 빈도

$\sum_k n_{k,j}$  : 문서  $d_j$ 에서 모든 단어 출현 빈도

문서  $d_j$ 에서 특정 단어  $t_i$ 가 나타나는 횟수를 문서  $d_j$  내의 모든 단어가 나타나는 횟수로 나누면  $tf_{i,j}$ 이 계산된다.

단어 자체가 전체 문서 내에서 자주 사용되는 경우 그 단어가 흔하게 등장한다는 것을 DF(Document Frequency)라고 하고, 이 값의 역수를 취한 것을 IDF(Inverse Document Frequency)라고 한다.<sup>[15]</sup> IDF는 해당 단어의 일반적인 중요도를 나타내는 수치이다. 아래 식을 보면  $|D|$ 는 전체 문서 숫자이며  $|\{d_j : t_i \in d_j\}|$ 은 특정 단어  $t_i$ 가 포함된 문서들의 숫자이다.<sup>[13]</sup> 전체 문서의 수를 특정 단어  $t_i$ 가 포함된 문서들의 수로 나눈 값을 로그로 취하면 아래의 식 (3)과 같이 표현된다.<sup>[16]</sup>

$$idf_i = \log \frac{|D|}{|\{d_j : t_i \in d_j\}|} \quad (3)$$

$|D|$  : 전체 문서의 수

$|\{d_j : t_i \in d_j\}|$  : 특정 단어  $t_i$ 가 포함된 문서의 수

만약, 특정 단어  $t_i$ 가 전체 문서에 존재하지 않을 경우 분모가 0이 되므로, 이를 예방하기 위해  $1 + |\{d_j : t_i \in d_j\}|$  로 사용되는 것이 일반적이나, 본 논문에서는 수집되는 개인정보만을 대상으로 하였기 때문에 원래의 IDF 식으로 계산 하였다.

TF-IDF 가중치 모델은 TF와 IDF의 곱으로 계산되고 아래 식(4) 으로 표현 된다.<sup>[15]</sup>

$$tf-idf_{ij} = tf_{ij} \times \log \left( \frac{N}{idf_i} \right) \quad (4)$$

따라서 본 논문에서는 개인정보의 위험도 측정을 위한 가중치를 계산하기 위해서 TF는 한 금융사의 개인정보별 수집 현황을 조사하고, IDF은 28개의 금융사를 대상으로 계산을 한다. 개인정보 가중치  $weight_{ij}$ 를 계산하여 이용한다.

## 2. IPA(Importance Performance Analysis)

### 기법

IPA(Importance Performance Analysis) 기법은 1977년에 Martilla & James가 Journal of Marketing에 자동차 사업의 성취도를 분석하기 위해 다속성 모델(Multi-attribute)을 기초로 한 Importance Performance Analysis라는 논문을 발표하면서 알려졌다.<sup>[17]</sup>

IPA기법은 중요도와 만족도를 동시에 분석하는 마케팅 기법으로서, 이용자가 중요하게 여기는 속성을 조사하고, 이후 수행도는 평가하여 개별 속성에 대하여 상대적인 중요도와 만족도를 비교 분석하는 평가 기법이다.<sup>[18]</sup> IPA기법의 사분면에 대한 정의 및 내용은 그림 1과 같다. 제1사분면은 매우 중요하지만 만족도가 낮은 속성들로 이루어져 관리자들이 집중해서 관리해야 한다. 제2사분면은 중요하고 만족도 역시 높은 속성들이 위치하게 되므로 관리자들은 현재의 수준을 유지하는 것이 중요하다. 제3사분면은 중요도와 만족도 모두 낮은 속성들로 평가 역량이 모두 낮은 상태로 상대적으로 개선 노력에 대한 필요성이 낮은 요소들이다. 제4사분면은 만족도가 높게 나타나지만, 중요도는 낮은 경우로 이 평가요소들에 대해서는 향후 시급히 개선해야 하는 속성이다.<sup>[19,20,21]</sup>

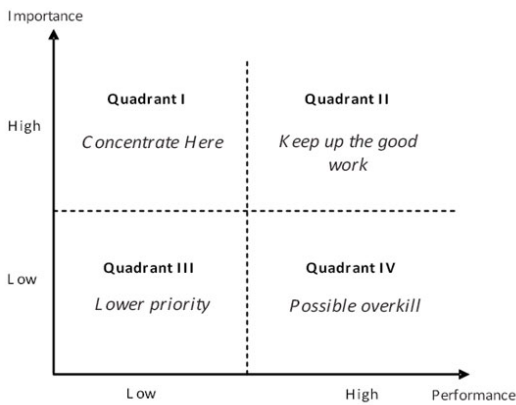


그림 1. 중요도-성과 분석 모델

Fig. 1. Importance-Performance Analysis Model

IPA기법에서 사분면의 영역을 구분하기 위해서는 중요도와 만족도의 중심점을 설정한다. 이는 위치에 따른 전략방안을 제시와 평가하는 항목의 위치 결정의 기준이 된다는 점에서 매우 중요하다. IPA를 사용한 연구들은 평균값을 중심축으로 채택·추천하고 있어, 본 연구에서도 평균값을 중심점으로 결정하여 IPA분석을 실시하였다.

## IV. 연구방법

본 논문에서 제안하는 방법의 목적을 효과적으로 수행하기 위해 개인정보수집 항목의 대상은 시중은행, 지방은행, 특수은행, 카드사 포함하여 총 28개의 금융회사를 중심으로 하였다. 각 금융사별 선택 항목 및 개인정보수집 항목 중 선택항목, 기타 금융 거래의 설정·유지·이행·관리를 위한 상담·채권관리 등을 통해 생성되는 정보를 제외한 필수 항목, 금융거래정보, 신용도 판단 정보를 수집하였다. 자료의 분석방법으로는 먼저, 개인정보 영향 평가의 개인정보 영향도 등급표를 참고 하여 중요도를 측정하였다. 금융권에서 개인정보 간의 상대적 가중치를 평가하기 위해 TF-IDF 기법을 통하여 측정하였다. 다음으로는 IPA기법을 적용하여 개인정보 위험도를 측정하였다. 아울러 수집된 데이터의 통계처리를 위해 SPSS 18.0 과 Micro Excel을 이용하여 통계치를 분석 하였다.

### 1. 개인정보 중요도 및 TF-IDF 가중치

개인정보 중요도는 개인정보 영향도 등급표를 통하여 개인정보 중요도 수치를 산출 하였다.

TF-IDF 가중치는 각 금융사들의 개인정보 취급방침을 통하여 알아낸 수집하고 있는 개인정보 개인 목록을 통하여 개수를 이용하여 TF-IDF 가중치를 계산 하였다.

표 1. 개인정보 중요도 및 가중치

Table 1. The importance and weight of personal information

Classification		importance	Collection number	Weight
General information	Name	3	28	3.33
	Resident registration number	5	28	3.33
	cell phone number	5	14	2.64
	Occupation	2	23	3.14
	Contacts	5	20	3.00
	Gender	1	3	1.10
Communication information	Nationality	1	24	3.18
	ID	3	3	1.10
	E-mail	4	24	3.18
Financial transaction information	Whether to receive E-mail	1	3	1.10
	The type of goods	2	22	3.09
	Interest rate collateral	2	22	3.09

	Transaction date and time	2	22	3.09
	amount	4	18	2.89
Credit capability information	Tax record	5	21	3.04
	workplace	2	7	1.95
	The total amount of assets and liabilities and income	5	16	2.77
Determining information	Subrogation	4	19	2.94
	Extreme Urgency	4	19	2.94
	Overdue	4	19	2.94
	Credit rating	3	3	1.10
	The fact of the relevant occurrence	4	18	2.89
Credit Information	Password preceding two digits	4	6	1.79
	Card number	5	5	1.61
	Card validity	5	4	1.39
	Financial order disordering Information	5	4	1.39
	CVC	5	3	1.10

## V. IPA 분석 결과

개인정보 위험도를 알아보기 위하여 IPA기법을 이용하여 조사한 결과는 그림 2와 같다. 금융권에서 수집하는 개인정보에 대한 중요도와 실행도를 2차원 공간 상에 일목요연하게 나타내었다. X축은 개인정보의 가중치를 나타내고 있으며, 우측으로 갈수록 그 정도가 높다. 또한 Y축은 개인정보 중요도를 의미하며 위로 올라갈수록 중요도가 높다. X축과 Y축의 접점은 각 요인 내 변수들의 평균값을 기준으로 설정 되었다.

X축인 가중치의 중앙값은 2.25 이며, Y축인 중요도 중앙값은 3.00 이다. 이를 이용하여 IP의 교차점을 설정하였고 X축은 개인정보 가중치를 Y축은 중요도를 이용하여 구상한 분석 결과를 4분면으로 표현하였다.

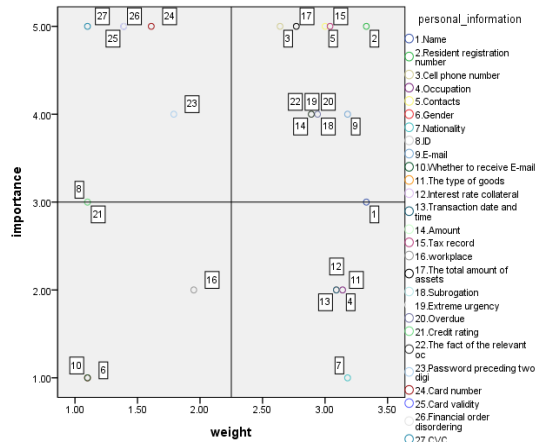


그림 2. 중요도-성과 분석 모델

Fig. 2. Importance-Performance Analysis Model

그림 2에서 나타난 것과 같이 개인정보 위험도 분석에서 제 I 사분면에는 8개, 제 II 사분면에 11개, 제 III 사분면에 2개, IV 사분면에 6개의 정보가 존재 하고 있다. 그리고 구간별 속성의 분포를 정리하면 표2와 같다.

표 2. 개인정보 중요도 및 가중치

Table 2. The importance and weight of personal information

Section	Property evaluation
quadrant I	CVC, Financial order disordering Information, Password preceding two digits, Credit quality, ID, workplace, Card number, Card validity
quadrant II	E-mail, The fact of the relevant occurrence, amount, Tax record, Subrogation, Large payment, Contacts, Overdue, cell phone number, The total amount of assets and liabilities and income, Resident registration number
quadrant III	Whether to receive E-mail, Gender
quadrant IV	Transaction date and time, Country of Citizenship, The type of goods, Name, Interest rate collateral, Occupation

그림 2와 표 2를 살펴 본 결과를 제 I 사분면은 중요도는 높으나, 가중치가 낮은 정보들로서, CVC, 금융질서문란정보, 비밀번호 앞 2자리, 신용도, 아이디, 직장, 카드번호, 카드유효기간은 매우 중요한 정보들이나 가중치가 낮아 자칫 개인정보 담당자들이 위험도를 낮게 측정하여 실무에 적용할 수도 있다.

제 II 사분면은 중요도와 가중치 모두 높은 개인정보들

로서 E-mail, 관련된 발생사실, 금액, 납세실적, 대위변제, 대지급, 연락처, 연체, 이동전화번호, 재산·채무·소득의 총액, 주민등록번호가 여기에 속한다.

제 3사분면은 중요도와 가중치 모두 낮은 개인정보들로서 E-mail 수신여부, 성별로 나타났다.

제4사분면은 중요도는 높으나 가중치가 낮은 항목들로서 거래일시, 국적, 상품종류, 성명, 이자율 담보, 직업으로 나타났다.

## VI. 결론

본 논문에서는 개인정보 중요도와 IF-IDF를 이용한 정량적인 가중치 분석을 통하여, 각 개인정보별 위험도를 산출해 내었다. 또한 IPA기법을 이용하여 분석하여 개인정보의 위험도를 정량적으로 산출해 내었다.

본 연구를 통해 사용자의 주관적인 판단을 배제할 수 있고, 결합된 개인정보 위험도 산정에 사용될 수 있으며, 정략적인 위험도를 통하여 객관적인 지표로 사용 될 수 있는 기준을 제시 하였다.

그러나 개인정보 수집 시 선택항목을 제외하여 개인정보를 수집한 부분이 한계점이다. 추후 연구에서는 이용자의 선택으로 수집되는 개인정보를 고려하여 평가하는 방법을 통한 종합적인 개인정보 위험도를 평가 한다 면 더 정확한 위험도를 산출 해 낼 수 있을 것이다. 또한 금융사뿐만 아니라 다른 분야에서도 개인정보 사고가 다수 일어나므로 업권별로 개인정보 수집을 하여 위험도를 산출하여 적용해 볼 수 있는 연구가 추가적으로 필요하다.

## References

- [1] Jaekwang Lee, Jongsu Jang, Kishik Park, "Personal Information Security in Cyberspace", The Korean Association For Information Society, Vol.12, 51-6, 2007.12
- [2] Woo-Jun Kang, "An Efficient Privacy Preserving Method based on Semantic Security Policy Enforcement", The Journal of The Institute of Internet, Broadcasting and Communication, VOL. 13, No. 6, 2013. 12
- [3] Daeseon Choi, Kim Seok Hyun, CHO, JIN-MAN, Seung-Hun Jin, "Big data privacy risks" Journal of the Korea Institute of Information Security and Cryptology, 23(6), p.56-p57
- [4] Kim Tae-hwan, Lee Ha-ney, yu jin-ho, "Subsequent spill of personal information and discussion of the variation pattern of stock prices of companies", The Korea Society of Management information Systems Spring Joint Scientific Meeting, pp.89-92, 2014.6
- [5] Korea Information Security Agency, "Guide to Information Security Management System Risk management", 2004.11
- [6] Privacy Information Protection Portal, "Privacy Impact Assessment education"
- [7] Korea Information Security Agency, "Privacy Impact Assessment education", 2011.12
- [8] Sung-kyu Cho, Moon-seog Jun, "Privacy Leakage Monitoring System Design for Privacy Protection", Journal of the Korea Institute of Information Security and Cryptology, 99-106, 2012.2
- [9] Ki-sung Lee, Hyo-beam Ahn, Su-Youn Lee, "A Study on a Risk Model to Protect Personal Information", Korea Information Assurance Society, Vol7. 2011.12
- [10] Myung-Ho Cheon, Jong-Seok Choi, Yong-Tae Shin, "Measuring method of personal information leaking risk factor to prevent leak of personal information in SNS", Journal of the Korea Institute of Information Security and Cryptology 23(6), 2013.12
- [11] Pyong Kim, Younho Lee, Timur Khudaybergenov, "A method for quantitative measuring the degree of damage by personal information leakage", Journal of The Korea Institute of Information Security & Cryptology VOL.25, NO.2, Apr. 2015
- [12] Sungjick Lee, Han-joon Kim, "Keyword Extraction from News Corpus using Modified TF-IDF", Journal of Society for e-Business Studies, 14(4):

59-73, 2009

- [13] Zhang, Yoshida, Tang, "A comparative study of TF-IDF, LSI and multiwords for text classification", Expert Systems with Applications 38, 2758 - 2765, 2011
- [14] wiki, <https://ko.wikipedia.org/wiki/TF-IDF>
- [15] NAJane, <http://blog.naver.com/choihk6067/120126938868>
- [16] Sungjick Lee, Han-joon Kim, "Keyword Extraction from News Corpus using Modified TF-IDF", The Journal of Society for e-Business Studies 14(4), 59-73, 2009.11,
- [17] J. A. Martilla and J. C. James, "Importance-Performance Analysis", Journal of Marketing, Vol. 41, pp. 77-79, 1977.
- [18] jo-seong lae, "A Study of the Aged in the Leisure Life of Leisure Motivation and on the Leisure Satisfaction", Honam for master's thesis. 2013. 2
- [19] Han Deok Han, Sun Ah Sung, Hee Yun Hwang, "A Study on Priority of residential environment factors according to the Urban-type Housing type", The Geographical Journal of Korea, vol48(2), 2014
- [20] Yeon-Sun Kim, Sang-Hee Lee, "Research on the Importance and Satisfaction of Selection Attribute for Pension using Importance-Performance Analysis(IPA)", The Journal of the Korea Contents Association, Volume 13, Issue 3, pp.392-401, 2013,
- [21] Yun, Yeo-Kyung · Won, Doyeon · Kwag, Min-Seok, "Evaluating service quality of Korean equestrian clubs using a revised IPA", The Korean Journal of Physical Education, 54(1), 315-326, 2015

## 저자 소개

### 정수진(준회원)



- 2009년 2월 : 동덕여대 컴퓨터과 (학사)
  - 2014년 3월 ~ 현재 : 고려대학교 정보보호대학원 석사 과정
- <관심분야 : 정보보호, 금융보안, 전자금융법규, 소프트웨어 테스트, 시큐어 코딩>

### 김인석(정회원)



- 1973년 : 홍익대학교 전자계산학과 (학사)
  - 2003년 : 동국대학교 정보보호학과 (석사)
  - 2008년 : 고려대학교 정보경영공학과 (박사)
  - 2009년 ~ 현재 : 고려대학교정보보호대학원 교수
- 現 FDS산업포럼 회장, 한국사이버정보전학회 운영위원등  
<관심분야 : 전자금융보안, IT 감사, 전자금융법규>

※ 본 연구는 미래창조과학부 및 한국인터넷진흥원의 "2015년 고용계약형 정보보호 석사과정 지원사업" 의 연구결과로 수행되었음