

# 이중 해쉬체인을 이용한 계층적 다중 처리를 위한 효율적인 데이터 관리 기법

정윤수\*, 김용태\*\*, 박길철\*\*

목원대학교 정보통신융합공학부\*, 한남대학교 멀티미디어학부\*\*

## An Efficient data management Scheme for Hierarchical Multi-processing using Double Hash Chain

Yoon-Su Jeong\*, Yong-Tae Kim\*\*, Gil-Cheol Park\*\*

Dept. of Information and Communication Convergence engineering, Mokwon University\*

Dept. of Multimedia Engineering, Hannam, University\*\*

**요약** 현재 인터넷을 통해 수집되는 빅 데이터는 데이터의 종류와 크기에 따라 데이터가 수집되는 시간보다 데이터가 증가하는 속도가 높아 사용자가 원하는 데이터를 원활하게 수집하는 것이 어려운 상황이다. 특히, 데이터의 사용 목적 및 종류에 따라 다르게 처리되기 때문에 데이터의 정확성과 계산비용이 빅 데이터 관리에 중요한 항목 중 하나이다. 본 논문에서는 인터넷에 존재하는 수많은 서로 다른 종류의 데이터를 사용자가 원할 때, 데이터를 정확하게 추출하는 동시에 데이터의 계산비용을 최소화하기 위해서 이중 해쉬체인을 이용한 계층적 다중처리 기반의 데이터 처리기법을 제안한다. 제안 기법은 다양한 종류의 데이터를 추출하기 위해서 데이터를 사용 목적 및 방법에 따라 계층적으로 분류한다. 이때, 데이터의 정확도를 높이기 위해서 데이터를 이중 해쉬체인으로 묶어 다중 처리한다. 또한, 제안 기법은 계층적으로 분류된 데이터를 손쉽게 접근하기 위해서 해쉬체인으로 데이터를 구성하여 데이터의 처리 비용을 줄였다. 실험결과, 제안 기법은 기존 기법보다 데이터의 정확도는 평균 7.8% 높았고, 데이터의 처리 비용은 4.9% 단축시켰다.

**주제어** : 빅 데이터, 데이터 관리, 이중 해쉬체인, 계층적 다중 처리, 데이터 정확성

**Abstract** Recently, bit data is difficult to easily collect the desired data because big data is collected via the Internet. Big data is higher than the rate at which the data type and the period of time for which data is collected depending on the size of data increases. In particular, since the data of all different by the intended use and the type of data processing accuracy and computational cost is one of the important items. In this paper, we propose data processing method using a dual-chain in a manner to minimize the computational cost of the data when data is correctly extracted at the same time a multi-layered process through the desired number of the user and different kinds of data on the Internet. The proposed scheme is classified into a hierarchical data in accordance with the intended use and method to extract various kinds of data. At this time, multi-processing and tie the data hash with the double chain to enhance the accuracy of the reading. In addition, the proposed method is to organize the data in the hash chain for easy access to the hierarchically classified data and reduced the cost of processing the data. Experimental results, the proposed method is the accuracy of the data on average 7.8% higher than conventional techniques, processing costs were reduced by 4.9% of the data.

**Key Words** : Big Data, Data Management, Double Hash Chain, Hierarchical Multi-processing, Data Accuracy

\* 이 논문은 2015년도 한남대학교 학술연구조성비 지원에 의하여 연구되었음

Received 19 August 2015, Revised 20 September 2015

Accepted 20 October 2015

Corresponding Author: Yong-Tae Kim(Hannam University)

Email: ky7762@hnu.kr

© The Society of Digital Policy & Management. All rights reserved. This is an open-access article distributed under the terms of the Creative Commons Attribution Non-Commercial License (<http://creativecommons.org/licenses/by-nc/3.0>), which permits unrestricted non-commercial use, distribution, and reproduction in any medium, provided the original work is properly cited.

## 1. 서론

최근 SNS(Social Network Service)와 페이스북(Facebook)과 같은 인터넷 서비스의 인기로 인하여 빅 데이터 서비스는 데이터뿐만 아니라 이미지, 소리, 동영상 등을 모두 포함하고 있다[1,2]. 빅 데이터는 데이터의 증가 속도 및 데이터의 형태와 질 측면에서 기존 서비스와 많은 다른 양상을 보이고 있기 때문에 기존 인터넷망에서 생성되고 서비스되는 데이터와는 많이 차별성이 존재한다[3,4,5].

소셜 네트워크 서비스를 통해 제공되는 데이터는 과거에 비해 데이터의 크기가 점점 작아지고 있다. 블로그나 SNS에서 유통되는 데이터는 사용자의 성향뿐만 아니라 소통하는 상대방의 관계 분석도 가능하게 되었다.

소셜 네트워크의 서비스가 다양화되면서 소셜 네트워크에서 사용되는 데이터의 양은 점점 증가하고 있고 데이터를 분석하는 전문적인 기술은 점점 더 요구되고 있는 상황이다[7,8]. 현재 소셜 네트워크에서 다루어지고 있는 빅 데이터는 정치, 사회, 경제, 문화, 과학 기술 등 거의 전 분야에서 사용되고 있으며 빅 데이터의 사용 목적에 따라 그 중요성은 더욱 부각되고 있다. 그러나, 데이터의 종류가 다양화되면서 데이터를 생성, 수집, 분석하는 과정에서 발생하는 통신량과 계산비용은 아직까지 문제로 남아있다[9,10].

소셜 네트워크에서는 빠른 검색속도도 중요하지만 데이터의 안전성도 많은 관심을 가지고 있다. 만약 수집된 데이터가 보안 문제로 유출된다면, 소셜 네트워크를 사용하는 모든 사람들의 정보가 유출될 수 있기 때문이다[4]. 빅 데이터와 관련된 최근 연구에서는 데이터를 효과적으로 분석하는데 초점을 두고 있다. 특히, 소셜 네트워크에서는 거의 실시간으로 데이터가 생성 및 삭제되기 때문에 데이터의 사용 목적 및 특성에 맞게 데이터를 관리하거나 데이터를 불법적으로 유출되는 것을 예방하는 연구가 진행중이다.

본 논문에서는 소셜 네트워크에서 처리되는 서로 다른 종류의 데이터를 저비용으로 정확하게 추출하기 위한 계층적 다중처리 기반의 데이터 처리기법을 제안한다. 제안 기법은 데이터의 특징 및 특성에 따라 데이터를 계층적으로 분류한 후 연관성이 있는 데이터를 해쉬 체인으로 묶어 데이터 추출시간을 단축시키는데 목적이 있다.

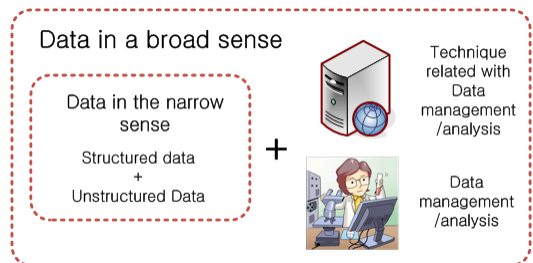
특히, 제안 기법은 수많은 데이터 중 유사 데이터의 분류 및 정확도를 높이기 위해서 단일 해쉬체인이 아닌 이중 해쉬체인으로 데이터 추출과정을 구성한다. 이 같은 과정을 수행하는 이유는 서로 다른 유형의 데이터를 종류, 기능, 특징에 따라 속성을 부여하고 데이터의 속성들에 따라 데이터를 서브넷으로 묶음으로써 여러 지역에 분산된 데이터를 손쉽게 연계함으로써 데이터의 접근성을 향상시키기 위해서이다.

이 논문의 구성은 다음과 같다. 2장에서는 빅데이터의 정의 및 데이터 추출 기법에 대해서 알아본다. 3장에서는 계층적 이중 해쉬 체인 기반의 데이터 추출 기법을 제안하고, 4장에서는 기존 데이터 추출 기법과 제안 기법을 비교 평가하고, 마지막으로 5장에서 결론을 맺는다.

## 2. 관련연구

### 2.1 빅데이터

빅 데이터는 기존의 관리 및 분석 체계로는 감당할 수 없을 정도의 거대한 데이터를 추출 및 분석하는 기술을 의미한다[1,9,10]. 빅 데이터는 정의 방법에 따라 좁은 의미의 데이터와 넓은 의미의 데이터로 구분되어 사용하고 있다. [Fig. 1]은 넓은 의미의 데이터와 좁은 의미의 데이터를 정의하고 있다.



[Fig. 1] Big Data Definition

[Fig. 1]처럼 좁은 의미의 데이터는 정형적 데이터와 비정형 데이터로 구분된다. 정형적 데이터는 데이터베이스에서 담을 수 있는 형태 데이터로 기업이나 기관의 데이터베이스에 저장되어 있거나 외부 데이터베이스의 데이터를 의미한다. 비정형 데이터는 언론사나 포털사이트에 뜬 뉴스 게시물, 블로그나, 온라인 커뮤니티 게시판의

게시물, 유튜브 등에 올라와 있는 동영상, 팟캐스트, 음악, 사진 등 다양한 데이터를 의미한다.

빅 데이터 정보는 민간 분야에서 뿐만아니라 다양한 분야(주요 도로와 공공건물, 아파트 엘리베이터 등)에서 만들어지고 있다[5,11]. 빅 데이터는 다양하고 방대한 규모의 데이터가 수십 테라바이트에서 수 페타바이트가 모여 특정 대상을 측정하고 경향을 예측하는 기술적 통계를 활용하는데 사용된다[12,13,14].

## 2.2 기존 연구

소셜 네트워크 서비스에 사용되는 데이터의 종류가 다양화되면서 데이터를 탐지하는 다양한 연구가 진행되고 있다. 특히, 최근 연구되고 있는 빅 데이터 탐지 알고리즘은 크게 트리기반 탐지 알고리즘, 해쉬 체인기반 탐지 알고리즘, 클러스터링 탐지 알고리즘 등으로 나뉜다 [6,7,8,15,16]. 트리기반 탐지 알고리즘은 개체를 거리, 속성에 따라 개체들을 트리형태로 구성하는 알고리즘이다. 트리 기반 탐지 알고리즘은 개체 탐지가 손쉬운 장점은 있지만 개체수가 증가할수록 개체 재배치 및 탐지 시간이 오래 걸리는 단점이 있다. 해쉬 체인기반 탐지 알고리즘은 개체를 해쉬함수에 적용하여 전체 개체를 연결하는 알고리즘으로써 개체를 손쉽게 구성할 수 있는 장점은 있지만 특정 개체를 찾을 경우 개체를 식별할 있는 정보가 추가로 필요한 단점이 있다. 해쉬 체인기반 탐지 알고리즘은 단점을 보완하기 위해서 클러스터링 탐지 알고리즘과 함께 사용하는 기법들의 연구가 증가하고 있다[6].

클러스터링 탐지 알고리즘은 계층형 클러스터링과 비계층형 클러스터링으로 구분된다. 계층형 클러스터링 기법은 개체들을 손쉽게 구성할 수 있지만 클러스터간 개체 이동이 어렵다는 문제점이 있다. 비계층형 클러스터링 기법은 계층형 클러스터링 기법에 비해 클러스터링에 소속된 개체의 이동이 손쉬워 클러스터를 재배치할 수 있는 장점이 있지만 클러스터상의 개체 이동시 업데이트 시간이 오래 걸린다는 문제점이 있다[7,8].

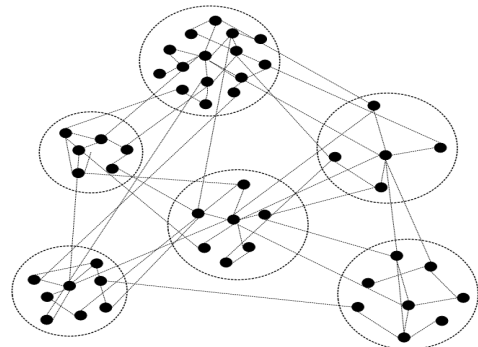
## 3. 계층적 해쉬체인 기반 데이터 추출 기법

이 절에서는 소셜 네트워크에서 등록되는 데이터 중

데이터의 양이 작은 정보들을 데이터의 종류, 기능, 특성에 따라 데이터를 계층적으로 분류하여 데이터간 유사도가 높은 데이터를 해쉬체인으로 묶음으로써 데이터의 정확도를 향상시킨다. 또한, 소셜 네트워크에 등록된 수많은 정보 중 사용자가 원하는 정보를 정확하게 추출하기 위해서 데이터 등록시 데이터를 선별할 수 있는 태그정보를 데이터와 쌍으로 사전에 서버에 등록한다. 여기서, 태그정보는 0과 1로 16진수 값이 나오도록 최소 32비트에서 최대 256비트로 구성한다고 가정한다.

## 3.1 개요

제안기법은 소셜 네트워크에 등록되는 다양한 종류의 데이터를 짧은 시간에 정확하게 검색할 수 있도록 데이터를 [Fig. 2]처럼 계층적으로 구성한다. 계층적으로 구성된 데이터는 데이터의 속성(날짜, 시간, 목적 등)에 따라 데이터와 속성 정보를 해쉬 체인으로 묶는다. 특히, 제안기법에서는 데이터의 연결정도를 백분율로 나타낸 데이터 연결 정도(DCI, Data Connection Information) 값을 해쉬체인의 시드값으로 사용기 때문에 데이터 재탐색 속도가 기존 해쉬체인 기법보다 높다.



[Fig. 2] Overall Structure of Proposed Scheme

[Fig. 2]처럼 계층적으로 구성된 데이터는 데이터의 연계 정보에 따라 계층적 구조가 수시로 변경될 수 있다. [Fig. 2]의 구조에서 데이터의 연계 정보는 식 (1)처럼 속성 값과 연결 정도값에 따라 구성된다. 식 (1)에서 행을 구성하는 정보는 속성 정보를 의미하고 열을 구성하는 정보는 연결 정도값을 의미한다. 식 (1)의 연결 정보는 [Fig. 2]처럼 데이터의 계층적 구조를 나타내는 척도로 사용된다.

$$C_k = \begin{pmatrix} x_{00} & \dots & x_{0j} \\ \dots & \dots & \dots \\ x_{i0} & \dots & x_{ij} \end{pmatrix}, k=1,2,\dots, n \quad \text{식 (1)}$$

여기서,  $i$ 는 데이터 속성 정보를 나타내는 행의 순서를 의미하며 범위는  $i=1,2,\dots, n$ 처럼 나타낸다.  $j$ 는 데이터간 연결 정도 값을 의미하며 범위는  $j=1,2,\dots, n$ 처럼 나타낸다. 데이터 연결 정도(DCI, Data Connection Information) 값은 데이터의 연결정도를 백분율로 나타내며 식 (2)처럼 표현한다.

$$DCI = \frac{P(IV)}{P(IV)P(C)} \quad \text{식 (2)}$$

여기서  $IV$ 는 데이터의 중요도를 의미하고,  $C$ 는 데이터 사용 횟수를 의미한다.

식 (2)의 연결 정도 값은 수 많은 주제로 구성된 데이터의 생성과 삭제에서 발생하는 데이터의 정확도에 영향을 미치며 사용자가 원하는 데이터를 빠르게 찾아 볼 수 있는 정보 탐색 시간을 정하는 항목으로 활용한다.

제안 기법은 소셜 네트워크와 같은 서비스에서 사용되고 있는 수 많은 데이터 중 사용자가 원하는 정보를 신속하면서도 정확하게 찾기 위한 연구에 초점을 갖고 있다. 그 중에서도 데이터의 속성 정보에 확률정보를 이용하여 데이터의 정확도에 따라 데이터의 구조가 계층적이면서 순차적인 구조가 갖도록 데이터를 동일 그룹으로 묶어 서브넷을 구성하도록 한다. 제안 기법은 이 같은 데이터 구성 및 관리 방법을 통해서 다양하고 복잡한 특성을 가지는 데이터를 기존 기법보다 정확하면서 계산비용을 줄이는 것을 목적으로 한다.

제안 기법에서는 사용자가 데이터를 검색하기 위해서 가장 먼저 데이터를 서버에 등록한다. 이때, 소셜 네트워크에 등록되는 데이터는 자동적으로 사용자가 작성한 데이터와 함께 데이터를 식별하기 위한 연계정보를 쌍으로 묶어 서버에 등록하게 된다. 서버에 등록된 데이터는 [Fig. 2]처럼 계층적으로 데이터를 구성하는 동시에 데이터의 분산처리가 가능하도록 한다. 이 때, 서버에 등록된 데이터는 연계정보의 확률값에 따라 데이터를 서브넷으로 묶어 데이터의 정확도를 향상시키기 위한 방법으로 활용한다.

### 3.2 용어 정의

<Table 1>은 제안 기법에서 사용하는 용어에 대한 설명이다.

<Table 1> Notations

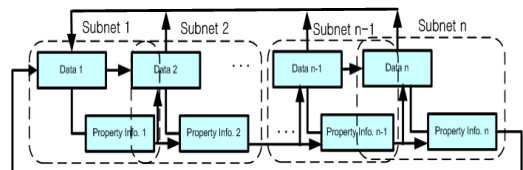
Notation	Definition
$i$	The order of rows representing the data attribute information
$j$	The degree of value between data connection
$CI$	Connection information of data stored within SNS
$DCI$	Data Connection Information
$P$	Probability of Data Connection Information
$IV$	Important value of data used within SNS
$C$	The number of usage of data

### 3.3 해쉬 체인 기반 계층적 구조의 데이터 추출 기법

이 절에서는 최근 소셜 네트워크에서 사용되는 데이터의 크기가 작아지면서 데이터의 관리가 어려운 상황을 해결하기 위해서 해쉬 체인 기반의 계층적 데이터의 추출기법을 제안한다. 제안기법에 사용되는 데이터 추출기법의 핵심은 데이터와 함께 데이터를 인식할 수 있는 연계정보를 데이터 속성정보(종류, 기능, 특성)에 따라 확률적으로 유사정도를 나타냄으로써 데이터를 확률값에 따라 서브넷으로 구성한다. 또한 서브넷으로 서로 연계된 데이터 정보는 다른 정보와 또다시 연계할 수 있도록 데이터 유사도 정보를 확률값으로 나타내도록 한다.

#### 3.3.1 데이터 처리 과정

소셜 네트워크에서 제공되는 수많은 데이터를 효율적으로 관리하기 위해서 제안 기법에서는 [Fig. 3]처럼  $n-1$  번째 데이터의 속성 정보를  $n$  번째 데이터와 함께 해쉬체인으로 연결하고 데이터와 속성 값은 각각 연계할 수 있도록 해쉬 체인으로 연결한다. 이 때, 데이터 간 연결은 확률값에 따라 변하며, 데이터 처리구조는 계층적 구조를 가지게 된다.



[Fig. 3] Overall Process Structure of Proposed Scheme

[Fig. 3]은 제안 기법에서 데이터간 연결 정도값을 구하기 위한 제안기법의 전체 처리 구조도를 보여주고 있다. [Fig. 3]처럼 제안 기법이 서버에 저장된 데이터를 데이터가 가지는 고유 정보 이외에 속성 정보를 반영한 것은 데이터가 가지고 있는 다양하고 복잡한 특성을 반영하여 데이터를 신속하게 처리하기 위해서이다.

제안 기법은 [Fig. 3]과 같은 처리 구조를 통해 데이터 속성에 확률값을 부여하여 데이터 유사도 확률에 따라 서브넷을 구성한다. 이 과정은 데이터의 정확도에 따라 수시로 변경되며 데이터의 계층적 구조도 매 순간 새로 생성하게 된다.

### 3.3.2 데이터간 연결 정보 처리 과정

제안 기법은 데이터를 속성 정보와 함께 계층적으로 확률값이 높은 데이터를 중심으로 서브넷을 구성한다. 제안 기법은 데이터 연결 정도를 처리하기 위해서 [Fig. 4]와 같은 알고리즘을 사용한다.

```

Hierarchical_Data(H, p)
1.  m = p.num_children_data
2.  k = m-1
3.  while k ≥ 0
4.      p.children[k].I = h(2k × p.I)
5.      Hierarchical_Data(H, p.children[k])
6.      k=k-1
    
```

[Fig. 4] Hierarchical Data Process Algorithm of Proposed Scheme

[Fig. 4]은 계층적 서브넷을 구성하기 위해서 해쉬 체인을 이용하였으며 계층적 해쉬체인을 표현하기 위해서 제안기법에서 사용된 데이터 연결 정보의 확률  $P$ 는 [16]을 기반으로 식 (4)처럼 나타낸다.

$$P = \begin{cases} \frac{n}{L} - \frac{n((L-1)!)^2}{L^2(L-2)!(L-1)!} & (L \geq 21) \\ \frac{n_a}{L} & (1 \leq L \leq 21) \end{cases} \quad \text{식 (4)}$$

여기서,  $L$ 은 해쉬체인의 길이를 의미하고  $n$ 은 계층적 데이터의 수를 의미한다.

식 (4)의 데이터 연결 확률값을 기반으로 제안기법에서는 확률값이 높은 데이터의 속성정보를 추출하여 정확도가 높은 데이터를 저비용으로 탐지하도록 서브넷을 생성한다.

## 4. 성능 평가

제안 기법의 성능평가는 데이터 수의 크기와 계산비용을 Wen-Guey 기법과 해쉬 트리 기법 등과 비교 평가한다.

### 4.1 환경설정

제안 기법의 성능 평가를 위해 제안 기법은 <Table 2>처럼 실험 환경을 설정한다. <Table 2>처럼 정확한 데이터를 검색하기 위한 서브넷의 수는 {1, 3, 5, 10}로 설정하고, 데이터 수는 {500, 1000, 2500, 5000, 7500, 10000}으로 설정하며, 데이터가 가지고 있는 속성수는 {1, 2, 3, 4, 5}로 설정한다.

<Table 2> Simulation Setting

Parameter	Setting
Number of Subnet	$s = \{1, 3, 5, 10\}$
Number of Selected Data	$d = \{500, 1000, 2500, 5000, 7500, 10000\}$
Number of Property	$p = \{1, 2, 3, 4, 5\}$

### 4.2 성능분석

#### 4.2.1 계층적 데이터 저장 공간 비교

<Table 3>는 소셜 네트워크에서 사용되는 데이터 관리 기법에서 가장 대표적으로 사용되고 있는 Wen-Guey 기법과 해쉬 트리 기법을 제안 기법과 서버의 데이터 저장 요구사항 크기를 비교하고 있다. <Table 3>에 사용되는 기법들은 모두  $N$ 개의 데이터를 가지고 완전 이중 트리와 깊이(height)  $h = \lceil \log N \rceil$ 으로 데이터를 구성하도록 가정하였다.

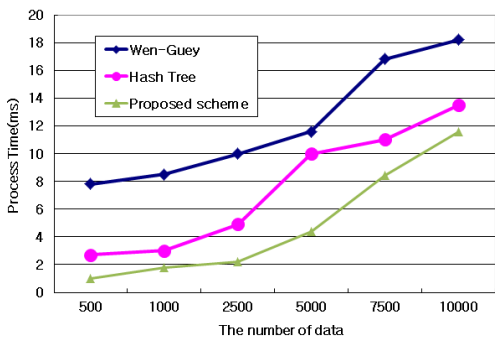
<Table 3> Hierarchical Data Storage Requirements

Technique	Wen-Guey	Hash Tree	Proposed Scheme
Data Storage Size	$O(N)$	$O(M \log N)$	$O(N)$

<Table 3>의 결과처럼 제안 기법은 Wen-Guey와 함께 데이터 저장 크기가  $O(N)$ 로써 차이를 보이지 않았지만, 해쉬 트리 기법의 저장 크기는  $O(M \log N)$ 로 제안기법보다 높게 나타났다.

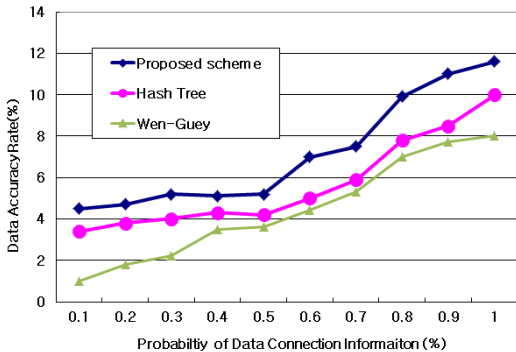
#### 4.2.2 계층적 데이터 처리시간

[Fig. 5]은 계층적으로 구성된 데이터 수에 따른 검색 처리 시간을 나타내고 있다. [Fig. 5]의 실험결과, 계층적으로 구성된 데이터 수가 증가할수록 해쉬 체인으로 구성된 데이터 속성정보의 확률값의 이용도는 높아져서 Wen-Guey 기법과 비교하였을 때 처리시간은 평균 6.5% 단축되었으며, 해쉬 트리 기법과 비교하였을 경우에는 처리시간이 8.5% 단축되었다. 이 같은 결과는 제안 기법에 사용되는 데이터 속성 정보에 따른 데이터의 연계 정보를 이용하였기 때문에 나타난 결과이다.



[Fig. 5] Process Time through the Number of Data

#### 4.2.3 확률 정보값에 따른 데이터 정확도



[Fig. 6] Data Accuracy Compare through Probability of Data Connection Information

[Fig. 6]은 확률 정보값에 따른 데이터의 정확도를 Wen-Guey 기법, 해쉬 트리 기법과 비교 평가한 결과이다. [Fig. 6]의 실험결과, 확률 정보값과 해쉬 함수를 이용

한 제안기법이 Wen-Guey 기법에 비해 데이터의 정확도가 평균 11.5% 높았으며, 해쉬 트리 기법에 비해 평균 13.1% 데이터의 정확도가 높게 나타났다. 이 같은 결과는 데이터와 속성 정보의 확률값을 쌍으로 묶어 계층적으로 구성한 후 데이터 확률값을 통해 데이터를 검색하였기 때문에 나타난 결과이다.

## 5. 결론

소셜 네트워크를 통해 수집되는 데이터는 종류와 크기에 따라 데이터의 처리시간과 활용방법이 다양해지고 있지만 데이터 수의 증가가 급격하게 증가하면서 정확한 데이터를 수집하는 것이 어려운 상황이다. 본 논문에서는 인터넷에 존재하는 수많은 서로 다른 종류의 데이터를 사용자가 원할 때, 데이터를 정확하게 추출하는 동시에 데이터의 계산비용을 최소화하기 위한 방법으로 해쉬 체인을 이용한 계층적 다중처리 기반의 데이터 처리기법을 제안하였다. 실험 결과, 데이터 처리시간은 Wen-Guey 기법과 비교하였을 때 처리시간은 평균 6.5% 단축되었으며, 해쉬 트리 기법과 비교하였을 경우에는 처리시간이 8.5% 단축되었다. 확률 정보값과 해쉬 함수를 이용한 제안기법이 Wen-Guey 기법에 비해 데이터의 정확도가 평균 11.5% 높았으며, 해쉬 트리 기법에 비해 평균 13.1% 데이터의 정확도가 높게 나타났다. 향후 연구로 본 연구의 결과를 기반으로 데이터 검색 시스템에 실제 적용할 계획이다.

## ACKNOWLEDGMENTS

This paper has been supported by 2015 Hannam University Research Fund.

## REFERENCES

- [1] H. Hu, Y. Wen, T. S. Chua, X. Li, "Toward Scalable Systems for Big Data Analytics: A Technology Tutorial", *IEEE Access*, vol. 2, pp. 652-687, 2014.
- [2] P. Russom, "Big Data Analytics", TDWI Research

- Fourth Quarter, pp. 6, Dec. 2011.
- [3] V. Gadepally, J. Kepner, "Big data dimensional analysis", 2014 IEEE High Performance Extreme Computing Conference(HPEC) pp. 1-6, Sep. 2014.
- [4] Y. Demchenko, C. De Laat, P. Membrey, "Defining architecture components of the Big data Ecosystem", 2014 International conference on Collaboration Technologies and Systems(CTS), pp.104-112, May, 2014.
- [5] J. Manyika, M. Chui, B. Brown, J. Bughin, R. Dobbs, C. Roxburgh, A. H. Byers, "Big Data: The Next Frontier for Innovation, Competition and Productivity", Mckinsey Global Institute, pp. 1-137. 2011.
- [6] P. Shen, Y. Zhou, K. Chen, "A Probability based Subnet Selection Method for Hot Event Detection in Sina Weibo Microblogging", 2013 IEEE/ACM International Conference on Advances in Social Networks Analysis and Mining, pp. 1410-1413, Aug. 2013.
- [7] K. Chen, Y. Zhou, H. Zha, J. He, P. Shen, X. Yang, "Cost-Effective Node Monitoring for Online Hot Event Detection in Sina Weibo", In Proceedings of the 22<sup>nd</sup> international conference on World Wide Web, ACM. pp. 107-108, April. 2013.
- [8] D. Kempe, J. Klenberg, E. Tardos, "Maximizing the spread of influence through a social network", In Proceedings of the ninth ACM SIGKDD international conference on Knowledge discovery and data mining, pp. 137-146, Aug. 2003.
- [9] K. M. P. Shrivastba, M. A. Rizvi, S. Singh, "Big Data Privacy Based on Differential Privacy a Hope for Big Data", 2014 International conference on Computational Intelligence and Communication Networks, pp. 776-781. Nov. 2014.
- [10] A. Katal, M. Wazid, R. H. Goudar, "Big data: Issues, challenges, tools and Good practices ", 2013 Sixth International Conference on Contemporary Computing(IC3), pp. 404-409, Aug. 2013.
- [11] Y. C. Jung. "Big Data revolution and media policy issues", KISDI Premium Report, Vol. 12, No. 2, pp. 1-22, 2012.
- [12] S. H. Kim, N. U. Kim, t. M. Chung, "Attribute Relationship Evaluation Methodology for Big Data Security", 2013 International Conference on IT Convergence and Security(ICITCS), pp. 1-4, Dec. 2013.
- [13] S. Y. Son, "Big data, online marketing and privacy protection", KISDI Premium Report, Vol. 13, No. 1, pp.1-26, 2013.
- [14] J. T. Kim, B. J. Oh, J. Y. Park, "Standard Trends for the BigData Technologies", 2013 Electronics and Telecommunications Trends, Vol. 28, No. 1, pp. 92-99, 2013.
- [15] M. Paryasto, A. Alamsyah, B. Rahardjo, Kuspriyanto, "Big-data security management issues", 2014 2<sup>nd</sup> International Conference on Information and Communication Technology(ICoICT), pp. 59-63, May, 2014.
- [16] Laurent, E. and V.D. Gligor, "A key-management scheme for distributed sensor networks". Proceedings of the 9th ACM Conference on Computer and Communications Security, ACM Press, pp. 41-47, Nov. 2002.

**정 윤 수(Jeong, Yoon Su)**



- 2000년 2월 : 충북대학교 대학원 전자계산학 이학석사
- 2008년 2월 : 충북대학교 대학원 전자계산학 박사
- 2009년 8월 ~ 2012년 2월 : 한남대학교 산업기술연구소 전임연구원
- 2012년 3월 ~ 현재 : 목원대학교 정보통신공학과 조교수

- 관심분야 : 센서 보안, 암호이론, 정보보호, Network Security, 이동통신보안
- E-Mail : bukmunro@mokwon.ac.kr

**김 용 태(Kim, Yong Tae)**



- 1984년 2월 : 한남대학교 계산통계학과 학사
- 1988년 2월 : 숭실대학교 전자계산학과 석사
- 2008년 2월 : 충북대학교 전자계산학과 박사
- 2002년 12월 ~ 2006년 2월 : (주)가림정보기술 이사
- 2010년 10월 ~ 현재 : 한남대학교 멀티미디어학부 교수
- 관심분야 : 모바일 웹서비스, 정보 보호, 센서 웹, 모바일 통신보안
- E-Mail : ky7762@hannam.ac.kr

**박 길 철(Park, Gil Cheol)**



- 1983년 2월 : 한남대학교 계산통계학과 학사
- 1986년 2월 : 숭실대학교 전자계산학과 석사
- 1998년 2월 : 성균관대학교 전자계산학과 박사
- 1998년 8월 ~ 현재 : 한남대학교 멀티미디어학부 교수
- 2005년 2월 : 한국정보기술학회 이사 멀티미디어 분과 위원장
- 관심분야 : Multimedia And Mobile Communication, Network Security
- E-Mail : gcpark@hnu.kr