

웹 정보자원의 내용기술을 위한 Keywords와 Description 메타태그 활용도에 관한 연구

A Study on the Use of Description and Keywords Meta Tags for the Content of WWW Resources

최재황(Jae-Hwang Choi)* · 조현양(Hyun-Yang Cho)**

< 목 차 >

- | | |
|--------------------------------|-------------------------------------|
| 1. 서론 | 2.2 메타태그의 분포 |
| 1.1 메타태그 | 2.3 Description 및 Keywords 메타태그의 분석 |
| 1.2 메타태그의 속성 | 2.4 Description 및 Keywords |
| 1.3 Description과 Keywords 메타태그 | 메타태그와 검색효율과의 관계 |
| 2. 연구방법 및 과정 | 3. 결론 및 향후 연구과제 |
| 2.1 데이터 추출 | |

초 록

본 연구의 목적은 웹 정보자원의 내용기술을 위하여 웹 상에서 어떠한 메타태그들이 어떻게 이용되고 있고, 가장 많이 이용되는 메타태그는 무엇이고, 이들과 검색효율과는 어떠한 관계를 가지고 있는가를 파악하는데 있다. 본 연구를 위하여 OCLC의 NetFirst에서 1,000개의 웹 문서가 추출·조사되었다. 1,000개의 웹 문서에서 총 92개의 메타태그를 발견하였고, 이중 “description”과 “keywords” 메타태그에 대하여 집중적인 분석이 이루어졌다. 또한, 먼저 검색된 웹 문서들은 나중에 검색된 웹 문서들보다 더 많은 “description”과 “keywords” 메타태그를 이용할 것이라는 가정아래 추출된 웹 문서들의 검색효율을 조사해 보았지만 그룹 간 큰 차이를 발견하지는 못하였다. 인터넷 일반검색엔진과 NetFirst와 같은 상용 DB와의 메타태그 이용에 대한 비교실험이 앞으로의 연구과제로 제시되었다.

키워드 : 메타태그, 메타데이터

Abstract

The purpose of this study is to investigate how and which meta tags are used, which meta tags are used frequently, and what relationships there are between retrieval of WWW documents and meta tags. For the study, 1,000 WWW documents were selected and examined from OCLC NetFirst. The total of 92 meta tags was discovered and “description” and “keywords” meta tags were analyzed intensively. In addition, Analysis of WWW documents showed that there are no significant relationships in meta tag usages between documents retrieved at the beginning and documents retrieved at the end. Comparative study between general internet search engines and commercial DBs such as NetFirst is suggested as a further study.

Key Words : description meta tag, keywords meta tag

* 한국과학기술정보연구원(KISTI) 정보유통사업실 선임연구원(findit@kisti.re.kr)

** 한국과학기술정보연구원(KISTI) 정보유통사업실 책임연구원(hycho@kisti.re.kr)

· 접수일 : 2001. 5. 15 · 최초심사일 : 2001. 5. 29 · 최종심사일 : 2001. 6. 18

1. 서론

검색엔진을 이용하여 웹 정보자원을 검색할 경우 찾고자하는 데이터에 정확히 접근하기란 쉬운 일이 아니다. 탐색대상이 되는 웹 정보자원의 수가 방대할 뿐만 아니라 해당 정보자원의 내용을 요약 기술하는 정보들이 너무나 복잡하고, 이용자들에게는 불필요한 요약정보들이 많이 포함되어 있기 때문이다. 웹 문서를 찾아 이들을 색인의 형태로 저장해 두는 프로그램의 일부는 HTML 파일에서 색인정보를 추출하기도 하는데 이때 이용되는 것이 메타태그이다.

HTML은 문서작성자에게 텍스트, 그래픽, 멀티미디어 자료 등의 물리적인 내용을 나타낼 수 있게도 하지만 메타태그를 이용하여 문서자체에 대한 정보도 제공할 기회를 준다. 메타태그는 키워드 검색엔진에서 웹 문서작성자들에게 그들의 문서가 웹 로봇에 의해 어떻게 색인될 것인지에 대해 통제를 가해줄 수 있는 수단이 되며, 올바른 형식과 잘 구성된 내용의 메타태그는 웹 문서의 검색효율을 높일 수 있다. Turner and Brackbill(1998)은 HTML 메타태그가 인터넷 검색엔진에서 정보검색을 향상시키는데 이용될 수 있다고 말하고 있다¹⁾.

본 연구는 웹 정보자원의 내용기술을 위하여 웹 문서들은 어떠한 메타태그들을 포함하고 있고, 웹상에서 주로 이용되는 메타태그들은 어떠한 것들이며, 메타태그는 어떠한 형식으로 이용되고 있으며, 가장 많이 이용되는 메타태그와 검색효율과는 어떠한 관계가 있는지를 파악하는데 있다. 본 연구를 위하여 1,000개의 웹 문서가 조사되었다.

1.1 메타태그

메타데이터는 일반적으로 “문헌에 대한 정보(information about a document)”, “정보자원에 대한 기술적인 정보(descriptive information about any information resource)”로 알려져 있으며 실제 콘텐츠는 아니면서 이에 대한 각종정보를 갖고 있는 데이터를 말한다. 의미상으로는 지금까지의 도서관 목록레코드와 다를 바가 없다. 그러나 전자문헌에 대한 메타 데이터의 개념은 웹 문서와 같은 네트워크 상의 정보자원을 쉽게 찾을 수 있도록 최소한의 핵심 데이터요소들만을 포함하고 있다는 것에 그 특징이 있다. W3C에서는 메타 데이터를 “웹 정보자

1) Turner, Thomas P. and Lise Brackbill(1998). “Rising to the Top: Evaluating the Use of the HTML META Tag to Improve Retrieval of World Wide Web Documents through Internet Search Engines”, *Library Resources and Technical Services*. 42(4). October 1998.

원에 대한 기계 가독정보(machine understandable information for the web)”로 정의를 내리고 있다²⁾.

메타태그는 메타데이터를 정의하기 위해 사용된다. 메타태그는 브라우저 상에는 나타나지 않으며 <TITLE> 태그와 마찬가지로 반드시 <HEAD>와 </HEAD>태그 사이에서만 사용되어지는 HTML 태그이며 마지막으로 </META>태그는 사용하지 않는다. HTML 명세(specification)에 의하면, 메타태그의 길이("<"에서 ">"까지의 길이)는 1,024바이트를 넘지 말아야 한다고 명시하고 있다.

1.2 메타태그의 속성

메타태그는 저자, 표제, 주제, 날짜, 언어, 키워드 등과 같이 웹 문서의 특성을 식별할 때 이용되며, 메타태그에 대한 값이 함께 따라 다닌다. 예를 들어, 웹의 저자를 명시하기 위해서 <표 1>의 <예 1>과 같은 메타태그가 이용될 수 있다.

<예 1>은 "Author"에 해당되는 메타태그를 명시하고 있으며, 여기에 "Dave Raggett"라는 메타태그 값을 부여해 주고 있다. 메타태그와 메타태그 값은 이중 따옴표("")로 둘러싸이게 된다. 여기서 표준으로 정해진 메타태그들의 목록은 없다. 메타태그와 메타태그에 대한 값은 프로파일(profile)이라 불리는 참고 사전에서 정의된다. 즉, 프로파일에는 "author", "copyright", "keywords" 등과 같은 특성이 정의된다.

메타 데이터의 장점을 저해하는 문제 중에 하나는 표준이 없다는 것이다. 이론적으로 무한의 메타태그를 만들어 낼 수도 있다. 그러나 누구도 이를 이용하지 않거나, 검색엔진 혹은 브라우저가 특별한 기능을 지원하지 않는다면 많은 메타태그는 사장되어 버릴 것이다. HTML 명세에 의하면 메타태그의 속성은 NAME, HTTP-EQUIV, CONTENT, SCHEME 속성으로 나누어진다³⁾.

NAME 속성<예 2>은 나타내고자 하는 메타태그의 이름을 표시한다. 앞에서 언급하였듯이 HTML 명세는 지원되어야 할 메타태그들의 목록을 정의하지는 않는다. NAME 속성은 "author", "document type", "distribution", "keywords", "description" 등의 여러 메타태그를 가질 수도 있지만, 메타태그의 이용을 지원하는 대부분의 인터넷 검색엔진들은 "keywords"와 "description"으로 정의된 NAME 속성만을 인식하고 이들을 색인에 이용하고 있다⁴⁾.

2) W3C Technology and Society. "Metadata and Resource Description", (<http://www.w3.org/Metadata>).

3) The global structure of an HTML document. "Meta Data", (<http://www.w3.org/TR/REC-html40/struct/global.html>)

4) Brenner, Diane and Marilyn Rowland(2000). *Beyond Book Indexing*. American Society of Indexers.

HTTP-EQUIV 속성<예 3>은 NAME 속성의 대신으로 사용될 수 있으며, http 헤더 특성(http header property)을 나타낼 때 이용된다. 즉, HTTP-EQUIV 속성은 HTTP 헤더와 동일하다. HTTP-EQUIV 속성과 NAME 속성의 차이점은 HTTP-EQUIV 속성은 http 헤더부분들을 나타내주는 반면 NAME 속성은 문서와 관련된 특성들을 더 반영한다는 것이다. NAME 속성과 마찬가지로 HTML 명세는 지원되어야 할 HTTP-EQUIV 속성들에 대한 특성들의 목록을 정의하지는 않는다⁵⁾.

CONTENT 속성<예 4>은 NAME 속성과 HTTP-EQUIV 속성에서 언급되었던 메타태그들의 값을 명시한다. 그리고 lang 속성이 CONTENT 속성에 대한 언어를 명시하는데 사용될 수 있다.

SCHEME 속성<예 5>은 CONTENT 속성에 나타난 값을 해석하거나 처리하기 위한 도구를 제시해 준다. 예를 들어, CONTENT 속성의 값으로 날짜를 "10-9-99"로 나타낼 수 있다. 이는 보통 사람에게 1999년 10월 9일인지 아니면 1999년 9월 10일인지 혼돈을 줄 수 있다. "Month-Day-Year"와 같은 SCHEME 속성의 값은 이 날짜의 값을 명확하게 해준다. <예 5>는 "0-8230-2355-9"의 값이 ISBN의 값을 나타내 준다. SCHEME 속성은 메타 데이터의 올바른 해석을 위한 문맥을 제공해 준다고 할 수 있다.

<표 1> 메타태그의 속성

<예 1>	<META name="Author" content="Dave Raggett"> (속성)(메타태그) (속성) (메타태그 값)
<예 2>	<META name="copyright" content="© 1997 Acme Corp.">
<예 3>	<META http-equiv="Expires" content="Tue, 20 Aug 1996 14:25:27 GMT">
<예 4>	<META name="author" lang="fr" content="Arnaud Le Hors">
<예 5>	<META scheme="ISBN" name="identifier" content="0-8230-2355-9">

1.3 Description과 Keywords 메타태그

문헌에 의하면, 가장 많이 사용되고, 검색효율과 관련 있는 메타태그는 "keywords"와 "description" 메타태그이다⁶⁾. "description" 메타태그는 웹 문서의 내용을 간략히 요약하는

Chapter 7. <Meta> Tags. pp.71-76.

5) Clark, Scott(1998). "META TAG TUTORIAL",
(http://www.webdeveloper.com/html/html_metatags.html)

부분이고, “keywords” 메타태그는 키워드를 나열하는 부분이다. 모든 검색엔진이 이 두 메타태그를 인식하는 것은 아니지만, 메타태그를 인식하는 검색엔진은 메타태그의 내용을 순위결정에 참조하므로 이의 효과적인 사용은 좋은 검색 결과를 가져 올 수도 있다⁷⁾. 그러나 Rowland(2000)는 “keywords” 및 “description” 메타태그가 인터넷 검색엔진 Hotbot과 infoseek에서는 검색결과에 영향을 미치기도 하지만, 이 두 메타태그를 수용하는 다른 검색엔진(예를 들면, AltaVista, Excite, WebCrawler 등)이 검색순위 랭킹에 꼭 영향을 미치는 것은 아니라고도 말한다⁸⁾. 본 연구는 실제로 “keywords” 메타태그와 “description” 메타태그가 가장 많이 이용되고 있는지 그리고 이 두 메타태그와 검색효율과는 어떠한 관계를 가지는지를 조사하게 된다.

“description” 메타태그는 해당 웹 문서의 내용을 요약해서 보여 준다. 일부 검색엔진은 검색결과를 출력할 때 웹 문서에 대한 설명을 위해 이 “description” 메타태그의 내용을 출력한다. “description” 메타태그에 대해서 표준으로 정해진 것은 없다. 그러나 일부 검색엔진들은 나름대로의 “description” 메타태그에 대한 기준을 가지고 있다. “description” 메타태그 정보를 지원하는 검색엔진에는 AltaVista, Hotbot, infoseek, Excite, WebCrawler 등이 있으며 “description” 메타태그의 길이는 150 바이트에서 395 바이트까지 검색엔진에 따라 다양하다⁹⁾. 그러나 대다수의 검색엔진은 infoseek이 정해놓은 200 바이트까지를 기준으로 따르고 있다¹⁰⁾.

“keywords” 메타태그를 수용하는 검색엔진들은 웹 문서의 내용에 대해 적용하는 각종 순위를 “keywords”의 내용에 적용하기도 하며 이 키워드의 내용은 순위에 많은 영향을 미치기도 한다. “keywords” 메타태그도 “description” 메타태그와 마찬가지로 정해진 표준은 없다. AltaVista, Hotbot, infoseek, WebCrawler 등이 “keywords” 메타태그를 지원하고 있으며, 특이한 것은 “description” 메타태그를 지원하는 Excite는 “keywords” 메타태그를 지원하지 않는다¹¹⁾. 일부 검색엔진들은 나름대로의 “keywords” 메타태그에 대한 기준을 가지고 있다. 한 연구에 의하면, 인터넷 검색엔진 AltaVista와 infoseek에서 “keywords” 메타태그만의 단독사용 및 “keywords”와 “description” 메타태그의 공동사용은 검색효율과 관련이 있었지만, “description” 메타태그의 단독사용은 검색효율과 관련이 없었다고 말하고 있다¹²⁾. <표 2>는

6) Search Engine Watch. “How to Use HTML Meta Tags”,
(<http://www.searchenginewatch.com/webmasters/meta.html>).

7) met · a · tags. (<http://www.metatags.net/>).

8) Brenner, Diane and Marilyn Rowland(2000). pp.71-72.

9) 최재황. “웹 정보자원의 색인과 초록 요소에 관한 연구”, 《한국문헌정보학회지》. 33(1). pp.113-128. 1999.

10) Craven, Timothy C. “Features of DESCRIPTION META tags in public home pages”, *Journal of Information Science* 26(5). pp.303-311, 2000.

11) Excite. “Understand Meta Tags”, (<http://www.excite.com/info/listing8.html>).

12) Turner and Brackbill(1998).

“description”과 “keywords” 메타태그가 HTML 태그 내에서 사용되는 일반적인 예이다.

<표 2> Description과 Keywords 메타태그의 사용 예

```
<HTML>
<HEAD>
  <TITLE>The Web Developer's Virtual Library</TITLE>
  <META NAME= "Keywords" CONTENT="HTML, CGI, Java,
    VRML, browsers, plugins, graphics, HTTP servers, JavaScript, Perl,
    ActiveX, Shockwave">
  <META NAME="Description" CONTENT=" Locate web
    authoring and software Internet resources at The WDWL,
    a well-organised goldmine with over 500 pages and thousands of
    links about HTML, CGI, Java, VRML, browsers, plugins, graphics,
    HTTP servers, JavaScript, Perl, ActiveX, Shockwave">
</HEAD>
```

2. 연구방법 및 과정

2.1 데이터 추출

본 연구에서 사용된 데이터는 한국교육학술정보원(<http://www.keris.or.kr/>)의 해외학술 DB 중 OCLC FirstSearch의 NetFirst를 이용하여 추출된 1,000건의 웹 문서로 구성되었다. 추출 방법은 2000년 12월 14일 NetFirst의 검색창에서 “library”, “information”, “science” 세 개의 무작위 단어를 입력하여, 여기서 검색된 3,913개의 웹 문서 중 처음 1,000개의 웹 사이트를 본 연구의 표본으로 정하고 이들 웹 문서의 메타태그들을 “소스보기”를 통하여 조사하였다.

이용된 “library”, “information”, “science” 세 개의 단어는 단지 원하는 수 만큼의 연구 표본을 얻기 위함이고 본 연구의 내용과는 관련이 없다. 그리고 1,000개의 웹 문서를 추출한 이유는 NetFirst가 최대 1,000개의 웹 문서만을 보여주고 있기 때문에 최대의 웹 문서를 뽑아 조사해 보기 위함이었다. 그리고 NetFirst를 선택한 이유는 NetFirst가 일반 검색엔진에서

찾을 수 있는 웹 문서들보다 정제된 자료를 수록하고 있기 때문에 질적인 면에서 나올 것이라는 것과 이들 웹 문서들이 메타태그를 더 많이 그리고 정확하게 이용하고 있을 것이라는 생각에서였다.

2.2 메타태그의 분포

1,000개의 웹 문서에서 뽑은 메타태그의 리스트는 <표 3>과 같다. <표 3>에서 보는 바와 같이 2회 이상 발생한 메타태그는 59개였고, 1회만 발생한 메타태그도 33개나 되었다. 이는 1,000개의 웹 문서에서 총 92(59+33)개의 메타태그가 이용되었음을 말해준다. 200회 이상 발생한 메타태그도 4개("content-type", "description", "generator", "keywords")나 되었다.

표본 데이터 가운데 가장 많이 발생한 메타태그는 "content-type"으로 250회였다. 한글과 같이 한 글자를 2바이트로 표현하는 국가에서는 일부 문서에서 문자가 알아볼 수 없는 이상한 기호로 나타나는 경우가 있다. 이때 문자 셋을 지정하면 문자가 제대로 나타나게 되는데, 이 역할을 "content-type" 메타태그가 하고 있다. 229회 발생한 "generator"는 해당 문서가 어떤 도구(예를 들면, Microsoft FrontPage 4.0)에 의하여 만들어 졌는지를 나타내 준다. "content-type"과 "generator" 메타태그는 사람이 만들었다고 할 수 없고 컴퓨터가 직접 생성하게 되며, 검색효율과는 관련이 없다. 200회 이상 발생한 일반 메타태그 중에서 검색효율을 염두에 두고 사람이 직접 만든 메타태그는 "description"과 "keywords" 메타태그 두 개 뿐이었다.

2.3 Description 및 Keywords 메타태그의 분석

200회 이상 이용된 메타태그에는 "description"과 "keywords" 메타태그 외에 "content-type"과 "generator"도 있다. 그러나 이들은 검색효율과 관련이 없고 브라우저 상에서 문자를 제대로 보여주는 역할을 하거나 컴퓨터가 자동으로 생성해 주기 때문에 이들에 대한 상세한 분석은 실시하지 않았다.

1,000개의 웹 문서 중에서 "description" 메타태그를 가지고 있었던 웹 문서는 206개(20.6%)였으며, "keywords" 메타태그는 229개(22.9%), 그리고 "keywords" 메타태그와 "description" 메타태그가 동시에 사용된 경우는 191개(19.1%)였다. 추출된 1,000개의 웹 문서 가운데 "description" 또는 "keywords" 메타태그를 포함한 웹 문서는 244개(24.4%)였다. 이는 1997년의 한 연구에서 밝힌 "keywords" 메타태그(12%)와 "description" 메타태그(11%)의 사용

<표 3> 1,000개의 웹 문서에서 나타난 메타태그와 그 예

Meta Tag 이 름	발생 횟수	예시
author	96	<meta name="Author" content="Chris Ferguson">
contact	2	<meta name="Contact" Content="www/dbk@nas.edu">
content-language	8	<meta http-equiv="Content-Language" content="en-us">
content-script-type	2	<meta http-equiv="Content-Script-Type" content="text/javascript">
content-style-type	2	<meta http-equiv="Content-Style-Type" content="text/css">
content-type	250	<meta http-equiv="Content-Type" content="text/html; charset=KS_C_5601-1987">
copyright	8	<meta name="COPYRIGHT" content="2000 CBS Information Technology Services">
created	4	<meta http-equiv="Created" content="08/07/00">
date	4	<meta name="Date" content="1999-05-17">
DC.Author	3	<meta name="DC.author" content="(TYPE=name)S.M.Finch">
DC.Contributor	4	<meta name="DC.Contributor" content="Michael Brundin">
DC.Creator	61	<meta name="DC.Creator" content="Tami Oliphant; Lisa Given">
DC.Creator.Address	4	<meta name="DC.Creator.Address" content="tmciolek@combs.anu.edu.au">
DC.Date	72	<meta name="DC.date" content="(scheme=ANSI.X3.30-1985)(type=created)19990220">
DC.Description	69	<meta name="DC.description" content="Students will fold paper to make an origami duck.">
DC.Form	2	<meta name="DC.form" content="(SCHEME=IMT) text/html">
DC.Format	64	<meta name="DC.format" content="(scheme=IMT)(type=contentType)text/html">
DC.Identifier	72	<meta name="DC.identifier" content="(scheme=GEM)(type=SID)AskERIC">
DC.Language	59	<meta name="DC.language" content="(SCHEME=ISO639) english">
DC.Package.Begin	56	<meta name="DC.package.begin" content="1">
DC.Package.End	56	<meta name="DC.package.end" content="1">
DC.Publisher	68	<meta name="DC.publisher" content="(type=role)onlineProvider">
DC.Relation	2	<meta name="DC.Relation" content="IsPartOf First Monday, vol. 5, no. 5">
DC.Rights	6	<meta name="DC.rights" content="(type=RL)http://www.askeric.org/Virtual/Lessons/copyright.html">
DC.Subject	76	<meta name="DC.subject" content="(scheme=GEM)(type=levelOne)Arts">
DC.Title	77	<meta name="DC.title" content="(lang=en)Origami Ducks">
DC.Type	72	<meta name="DC.type" content="(scheme=GEM)Lesson plan">
description	206	앞의 예문 참조
distribution	7	<meta name="distribution" content="global">
expires	7	<meta name="expires" content="Tuesday, 6-Oct-98 05:00:00 GMT">
formatter	2	<meta name="FORMATTER" content="Microsoft FrontPage 2.0">
GEM.Audience	54	<meta name="GEM.audience" content="(scheme=GEM)(type=beneficiary)Students">
GEM.Cataloging	56	<meta name="GEM.cataloging" content="(type=application)GEMCat">
GEM.Grade	56	<meta name="GEM.grade" content="(scheme=GEM)(type=grade)2,3,4">
GEM.Package.Begin	56	<meta name="GEM.package.begin" content="1">
GEM.Package.End	56	<meta name="GEM.package.end" content="1">
generator	229	<meta name="GENERATOR" content="Microsoft FrontPage 4.0">
GroupStart	56	<meta name="GroupStart" content="6">
GroupEnd	56	<meta name="GroupEnd" content="6">
keywords	229	앞의 예문 참조
keyword phrases	2	<meta name="keyword phrases" content="library and information science, computer science, University of South Florida School of Library and Information Science">
last-modified	4	<meta http-equiv="Last-Modified" content="Monday, 07 August, 2000">
language	3	<meta name="language" content="English">
Microsoft Border	16	<meta name="Microsoft Border" content="tb">
Microsoft Theme	12	<meta name="Microsoft Theme" content="school2000 010, default">
owner	4	<meta Name="Owner" content="WMO">
package	2	<meta name="package" content="(TYPE=begin) Dublin Core">
PICS-label	7	<meta http-equiv="PICS-Label" content="(PICS-1.1 "http://www.rsac.org/ratingsv01.html" 1 gen true comment "RSACi North America Server" by "billb@eskimo.com" for "http://www.amasci.com/" on "1997.10.30T09:47-0800" r (n o s o v o l o))">
post info	2	<meta name="post info" content="/scripts/post info.asp">
pragma		<meta name="Pragma" content="no-cache">
ProgId	19	<meta name="ProgId" content="FrontPage.Editor.Document">
publisher	5	<meta name="publisher" content="School of Library and Information Studies,University of Alberta">

Meta Tag 이름	발생 횟수	예시
refresh	7	<meta http-equiv="refresh" content="12; URL=http://www.aussi.org">
resource type	6	<meta name="resource-type" content="document">
reply-to	7	<meta http-equiv="Reply-To" content="fatomei@hotmail.com">
robots	18	<meta name="robots" content="index, nofollow">
subject	4	<meta name="subject" content="academic scholarships, aid university scholarship, work study">
title		<meta name="title" content="School of Library and Information Studies,University of Alberta">
usf	3	<meta name="usf" content="austind@chumal.cas.usf.edu">
발생횟수가 1인 메타태그		categories, classification, content-script, country, custodian, DC.CreatorContact, DC.Creator.email, DC.Creator.name, DC.Creator.organization, DC.Date.Y, DC.Keywords, DC.Publisher.Address, DC.Relation.IsVersionOf, DC.Relation.IsParentOf, DC.Source.definition, designed, editor, GEN.pedagogy, guide, Identifier-URL, key phrases, location, modified, object, organization, originator, packaging, postinfo, revised, source, version, window-target

통계보다 2배나 높은 수치이다¹³⁾. 결과적으로 1,000개의 웹 문서에서 “keywords” 메타태그를 “description” 메타태그보다 2.3%(23개) 더 많이 이용한 것으로 조사되었고, 이는 “keywords” 메타태그의 이용이 “description” 메타태그의 이용보다 1% 높았던 위의 연구[Sullivan(1997)의 연구]와도 동일한 경향이었다.

“keywords” 및 “description” 메타태그에 대한 상세한 분석을 위하여 “META MEDIC 3.2” META TAG Tester¹⁴⁾가 이용되었다. 분석은 두 메타태그 각각 다음의 항목을 중심으로 이루어졌다.

- 전체(total) 평균 메타태그 사이즈
- 실제(actual) 평균 메타태그 사이즈(순수한 content 속성 안의 내용에 대한 메타태그 사이즈)
- 실제 평균 메타태그 사이즈의 분포
- 메타태그의 정상적인 사용 여부

(1) Keywords 메타태그 분석

“전체 평균 keywords 메타태그”의 사이즈는 291.4 바이트였으며, “실제 평균 keywords 메타태그 사이즈”는 255.0 바이트였다. HTML 명세에서 추천하고 있는 “keywords” 메타태그의 사이즈는 없지만 “keywords” 메타태그를 지원하는 검색엔진들은 “실제 keywords 메타태그의 사이즈”로 300에서 500 바이트를 추천하고 있다. 그러나 본 연구에서 “실제 keywords 메타태그 사이즈”가 300-500 바이트 이내인 것은 229개의 “keywords” 메타태그 중에서 29개

13) Sullivan, Danny(1997), “The New Meta Tags Are Coming - Or Are They?”, (<http://www.searchenginewatch.com/sereport/97/12-metatags.html>).

14) EONS, “META Tag Tester: META Medic”, (<http://eons.org>).

(12.7%)에 불과하였다. “keywords” 메타태그의 “실제 keywords 메타태그 사이즈”의 분포도는 <표 4>와 같으며 0-300 바이트 사이가 229개의 메타태그 중에서 177개로 77.3%를 차지하고 있었다. HTML에서 제한하고 있는 1,024 바이트가 넘는 것도 8개나 되었다.

<표 4> 실제 “keywords” 메타태그 사이즈 분포

000-050	051-100	101-150	151-200	201-250	251-300	301-350	351-400	401-450	451-500	501-600	601-700	701-800	801-900	901-1000	1001-2000	2001-	계
26	45	34	35	18	19	10	4	9	6	5	3	4	0	3	6	2	229

229개의 “keywords” 메타태그를 분석해 본 결과 특이한 경우는 아래와 같다.

첫째, “keywords”라는 메타태그 이름이 조사결과 매우 다양한 것으로 나타났다. 유형으로는 “keyword”, “KEYWORD”, “keywords”, “KEYWORDS”, “KeyWord”, “KeyWords” 등이었다. “DC.Subject(76회 발생)”와 “key phrases(2회 발생)”메타태그는 “keywords” 메타태그에서 제외하였다.

둘째, 메타태그를 잘못 이용한 경우가 있었다. 태그 마지막에 “>”를 빼먹은 경우가 두 번<예 6>, 태그의 마지막에 “/”를 첨가한 경우가 한 번<예 7>, 그리고 CONTENT 속성의 내용에서 이중따옴표를 빠뜨린 경우가 한 번 있었다<예 8>.

셋째, “keywords” 메타태그를 중복 사용한 경우가 세 번 있었다<예 9>.

넷째, 보통 name 속성이 먼저 오고, content 속성이 뒤따르는 것이 일반적이지만, content 속성이 먼저 오고 name 속성이 뒤따르는 경우이다. 이와 같은 경우가 다섯 번 있었다<예 10>.

다섯째, CONTENT 속성의 이중따옴표 안에 내용이 아무 것도 없는 경우이다. 태그만 존재한 경우이고, 이와 같은 경우가 다섯 번 있었다<예 11>.

<표 5> “keywords” 메타태그의 분석결과 특이한 경우의 예

<예 6>	<META name="keywords" content="library, museum"
<예 7>	<META name="keywords" content="library, museum"/>
<예 8>	<META name="keywords" content=library, museum>
<예 9>	<META name="keywords" content="library, museum"> <META name="keywords" content="library, museum">
<예 10>	<META content="library, museum" name="keywords">
<예 11>	<META name="keywords" content="">
<예 12>	<META name="keywords" lang="fr" content="Arnaud Le Hors">

여섯째, LANG 속성이 이용된 경우이고, 229개의 “keywords” 메타태그 중에서 단 한 개의 웹 문서에서만 LANG 속성이 추가되었다<예 12>.

(2) Description 메타태그 분석

“전체 평균 description 메타태그”의 사이즈는 202.3 바이트였으며, “실제 평균 description 메타태그 사이즈”는 164.9 바이트였다. META MEDIC 태그 테스트, infoseek 및 대다수의 검색엔진이 정하고 있는 200 바이트에 비하면 평균 35 바이트 정도 모자라는 것이다. “실제 description 메타태그 사이즈”가 200 바이트 미만인 것은 206개의 “description” 메타태그 중에서 155개로 75.2%를 차지하였다. 1,024 바이트가 넘는 것은 한 개에 불과했다. “description” 메타태그의 “실제 description 메타태그 사이즈”의 분포도는 <표 6>과 같다.

<표 6> 실제 “description” 메타태그 사이즈 분포

000-050	051-100	101-150	151-200	201-250	251-300	301-350	351-400	401-450	451-500	501-600	601-700	701-800	801-900	901-1000	1001-2000	2001-	계
37	47	36	35	10	7	9	7	4	3	8	1	1	0	0	1	0	206

206개의 “description” 메타태그를 분석해 본 결과 특이한 경우는 아래와 같으며 “keywords” 메타태그와 매우 유사하였다. 이는 80%에 해당하는 웹 문서들이 “keywords” 메타태그와 “description” 메타태그를 동시에 이용했기 때문이었다.

<표 7> “description” 메타태그의 분석결과 특이한 경우의 예

<예 13>	<META name="description" content="library is a growing organism"
<예 14>	<META name="description" content="library is a growing organism"/>
<예 15>	<META http-equip="description" content="library is a growing organism">
<예 16>	<META content="library is a growing organism" name="description" >
<예 17>	<META name="description" content="">
<예 18>	<META name="description" lang="en-us" content="The Virtual Hospital is a continuously updated digital health sciences library. Its mission is to provide rapid, convenient access to high-quality medical information for health care providers and patients.">

첫째, “description”이라는 메타태그 이름이 여러 형태로 나타났다. 유형으로는 “description”, “Description”, “DESCRIPTION” 등이었다. “DC.DESCRPTION”(69회 발생) 메타태그는 “description” 메타태그에서 제외하였다.

둘째, 메타태그를 잘못 이용한 경우이다. 태그 마지막에 “>”를 빼먹은 경우<예 13>와 태그마지막에 “/”를 첨가한 경우<예 14>가 각각 한 번씩 있었다.

셋째, NAME 속성 대신에 HTTP-EQUIP 속성이 이용된 경우이다. 이러한 경우가 두 번 있었다<예 15>.

넷째, NAME 속성과 CONTENT 속성의 위치가 바뀐 경우이고, 이와 같은 경우가 네 번 있었다<예 16>.

다섯째, CONTENT 속성에 내용이 없는 경우이며, 이와 같은 경우가 네 번 있었다<예 17>.

여섯째, LANG 속성이 이용된 경우이고, 1개의 웹 문서에서 LANG 속성이 이용되었다<예 18>.

2.4 Description 및 Keywords 메타태그와 검색효율과의 관계

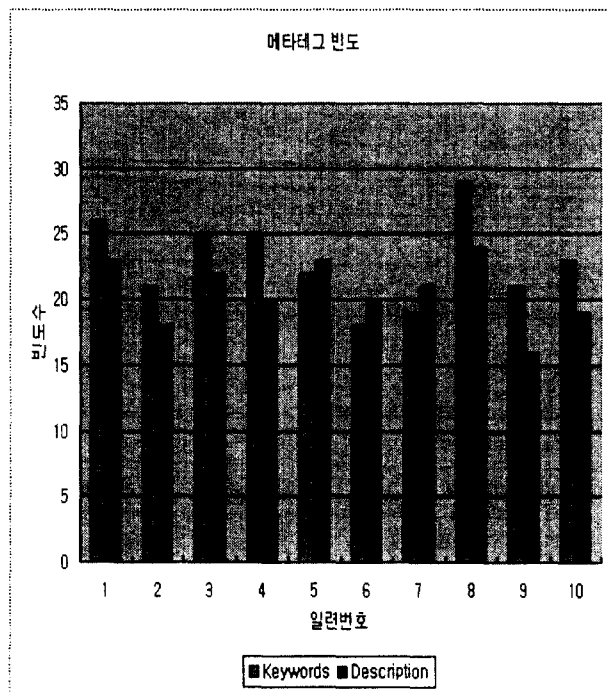
가장 많이 이용되는 메타태그와 이들을 사용한 웹 문서는 검색효율과 어떤 관계를 가지는지 그리고 먼저 검색된 웹 문서들은 나중에 검색된 웹 문서들보다 더 많은 메타태그를 이용하고 있는지에 대한 조사가 이루어졌다. 이를 위하여 1,000개의 웹 문서들을 100개 단위로 하여 10개로 나누고 각각의 그룹에 포함된 “keywords”와 “description” 메타태그의 개수를 확인하였다.

결과는 <표 8> 및 <그림 1>에서 보는 바와 같이, 처음에 검색된 웹 문서와 나중에 검색된 웹 문서 사이에서 메타태그 이용 빈도에 큰 차이가 없었다. 100개씩 10개의 묶음에서 두 메타태그는 각각 7-12%(16개-29개)로 골고루 나타났기 때문이다. “keywords”와 “description” 메타태그를 함께 사용한 웹 문서에 대하여도 조사가 이루어졌지만 <표 9>에서 보는 바와 같이 처음에 검색된 웹 문서와 나중에 검색된 웹 문서의 그룹 간 차이는 크지 않았다. 즉, OCLC의 NetFirst에서 메타태그의 사용과 검색효율과는 관련이 없는 것으로 조사되었다. 이는 크게 다음의 두 가지 이유로 해석되었다.

첫째, 로봇 프로그램을 이용하는 일반 검색엔진(예를 들면, AltaVista, Hotbot, infoseek, Excite 등)과 사람의 손을 거치는 NetFirst와는 차이가 있었기 때문이다. 즉, 1,000개의 웹 문서를 뽑은 NetFirst는 키워드 검색엔진이라기 보다는, 다시 말하면 로봇에 의존한다기 보다는, 사람의 노력이 들어간 디렉토리형 검색엔진에 가까웠기 때문이었다.

<표 8> keywords 및 description 메타태그의 빈도 및 퍼센트

일련번호	메타태그 빈도		메타태그 %	
	Keywords	Description	Keywords	Description
1 - 100	26	23	11.4	11.2
101 - 200	21	18	9.2	8.7
201 - 300	25	22	10.9	10.7
301 - 400	25	20	10.9	9.7
401 - 500	22	23	9.6	11.2
501 - 600	18	20	7.8	9.7
601 - 700	19	21	8.3	10.2
701 - 800	29	24	12.7	11.6
801 - 900	21	16	9.2	7.8
901 - 1000	23	19	10.0	9.2
계	229	206	100.0	100.0



<그림 1> keywords 및 description 메타태그의 빈도

둘째, 검색결과물의 랭킹이 검색엔진의 검색 랭킹 알고리즘에 따라 서로 상이하기 때문이다. 검색결과물의 랭킹은 검색엔진의 랭킹 알고리즘에 의해 정해지는 것이 일반적이다. 검색엔진에 따라 <title> HTML 태그에서 키워드를 발견했을 때, “keywords” 또는 “description” 메타태그에서 키워드를 발견했을 때, 또는 본문에서 키워드를 발견했을 때 등 부여하는 랭킹 알고리즘이 모두 다르기 때문이다. 예를 들어, “keywords” 또는 “description” 메타태그에서 키워드를 발견했을 때 점수를 5점을 주고 본문에서 키워드가 발견되었을 때 1점을 주더라도, 본문에서 10번의 해당 키워드가 발생하면 메타태그의 사용에 관계없이 검색 랭킹이 바뀔 수도 있기 때문이다. NetFirst의 검색엔진 담당자에게 검색랭킹의 순위에 대해 자문을 구해 보았으나 보안상 제공이 불가능하다는 답변을 받았다.

<표 9> keywords 및 description 메타태그가 모두 있는 경우의 빈도 및 퍼센트

일련번호	Keywords와 Description 태그가 모두 있는 경우	Keywords 태그만 있는 경우	Description 태그만 있는 경우
1-100	22	4	1
101-200	15	6	3
201-300	20	5	2
301-400	20	5	0
401-500	20	2	3
501-600	17	1	3
601-700	19	0	2
701-800	24	5	0
801-900	15	6	1
901-1000	19	4	0
계	191(전체태그의 87.8%)	39(전체태그의 9.8%)	15(전체태그의 3.4%)

3. 결론 및 향후 연구과제

웹 문서들은 메타태그를 어떻게 이용하고 있으며, 이용하고 있는 메타태그와 검색효율과는 어떠한 관계가 있을까라는 의문을 가지고 본 연구는 시작되었다. 이를 위하여 1,000개의 웹 문서들이 조사되었으며, 여기에서 56개의 일반 메타태그와 36개의 더블링크어 메타태그, 총 92개의 메타태그를 발견하였다. 92개의 메타태그 중 200회 이상 사용된 메타태그는 “content-type”(250회), “generator”(229회), “keywords”(229회), “description”(206회) 4개였지만, 검색효율과 관련이 있는 두 개(“description”과 “keywords”)의 메타태그에 대하여만 심층적으로 분석되었다.

1,000개의 웹 문서 중에서 “description” 메타태그를 가지고 있었던 웹 문서는 206개(20.6%)였으며, “keywords” 메타태그는 229개(22.9%), 그리고 “keywords” 메타태그와 “description” 메타태그가 동시에 사용된 경우는 191개(19.1%)였다. 추출된 1,000개의 웹 문서 가운데 “description” 또는 “keywords” 메타태그를 하나라도 포함한 웹 문서는 244개(24.4%)였다. 이는 244개의 웹 문서 중 약 80%에서 두 메타태그가 동시에 사용되고 있음도 나타내고 있다. 향후 인터넷 검색엔진은 “keywords” 및 “description” 메타태그에 대한 검색을 우선하고, 실제 검색 결과와 랭킹에 영향을 끼치게 되어 이들 메타태그의 활용은 앞으로도 더욱 증가하게 될 것이다.

“description” 및 “keywords” 메타태그를 사용한 웹 문서와 검색효율과의 관계에 대한 조사도 함께 이루어졌지만 웹 문서를 추출한 NetFirst에서 두 메타태그의 사용과 검색효율과는 관련이 없는 것으로 조사되었다. 이는 로봇 프로그램을 이용하는 일반 검색엔진과 사람의 손을 거치는 NetFirst와는 차이가 있었고, 검색결과물의 랭킹이 검색엔진의 검색 랭킹 알고리즘에 따라 서로 상이하였기 때문으로 풀이되었다.

본 논문이 보다 설득력을 갖기 위해서는 웹 DB별 메타태그 이용에 관한 비교연구가 수행되어야 할 것이다. 본문에서 언급한 메타태그를 수용하는 검색엔진(예를 들면, AltaVista, Hotbot, infoseek 등)과 메타태그를 수용하지 않는 검색엔진, 그리고 NetFirst와 유사한 상용 웹 DB간 메타태그 이용 및 검색효율이 비교 조사되어야 할 것이며, 검색엔진별 검색결과물 랭킹 알고리즘도 함께 비교 분석되어야 할 것이다.

참 고 문 헌

- 최재황(1999). “웹 정보자원의 색인과 초록 요소에 관한 연구”, 《한국문헌정보학회지》. 33(1). pp.113-128.
- Brenner, Diane and Marilyn Rowland(2000). *Beyond Book Indexing*. American Society of Indexers. Chapter 7. <Meta> Tags. pp.71-76.
- Clark, Scott(1998). “META TAG TUTORIAL”,
(http://www.webdeveliper.com/html/html_metatags.html)
- Craven, Timothy C(2000). “Features of DESCRIPTION META tags in public home pages”, *Journal of Information Science*. 26(5). pp.303-311.
- Excite(2000). “Understand Meta Tags”, (<http://www.excite.com/info/listing8.html>).
- EONS(2001). “META Tag Tester: META Medic”, (<http://eons.org>).
- The global structure of an HTML document(2000). “Meta Data”,
(<http://www.w3.org/TR/REC-html40/struct/global.html>).
- met · a · tags(2000). (<http://www.metatags.net/>).
- Search Engine Watch. “How to Use HTML Meta Tags”,
(<http://www.searchenginewatch.com/webmasters/meta.html>).
- Sullivan, Danny(1997). “The New Meta Tags Are Coming - Or Are They?”,
(<http://www.searchenginewatch.com/sereport/97/12- metatags.html>).
- Turner, Thomas P. and Lise Brackbill(1998). “Rising to the Top: Evaluating the Use of the HTML META Tag to Improve Retrieval of World Wide Web Documents through Internet Search Engines,” *Library Resources and Technical Services*. 42(4). October 1998.
- W3C Technology and Society. “Metadata and Resource Description”,
(<http://www.w3.org/Metadata>).