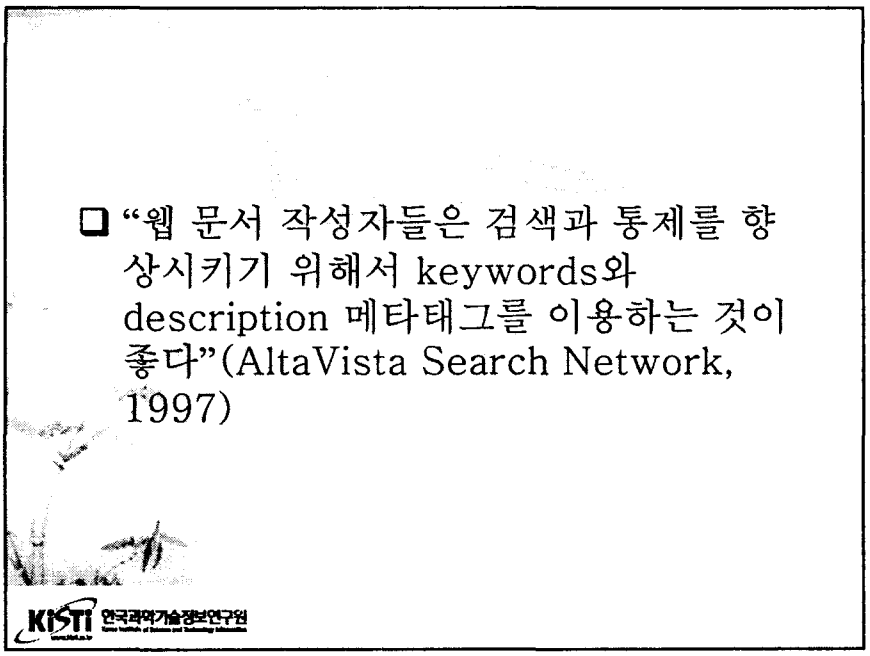
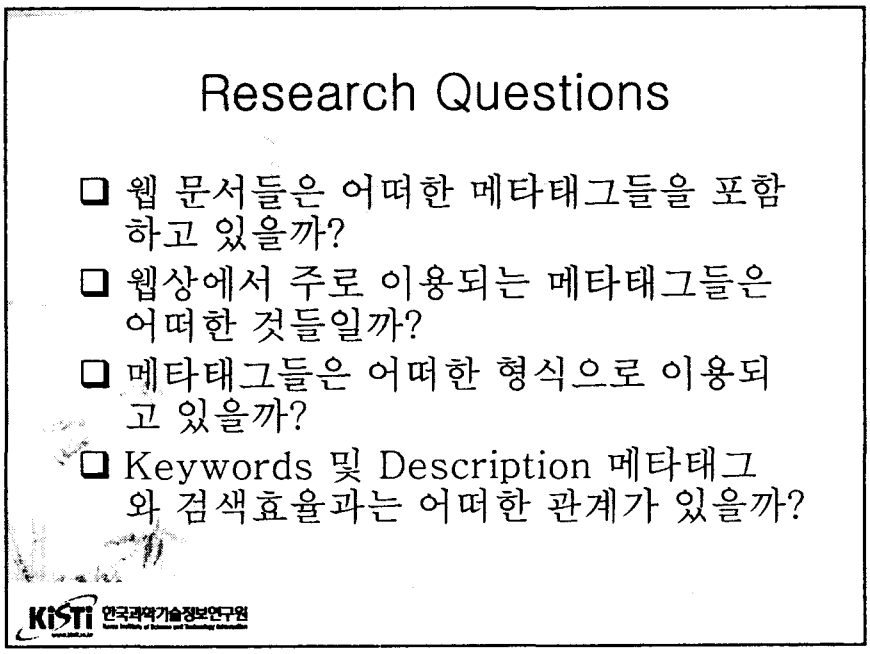
Keywords와 Description 메타태그 활용도에 관한 고찰

2001. 6. 29.

최 재 황
한국과학기술정보연구원 책임연구원

- "HTML 메타태그가 인터넷 검색엔진에서 정보검색을 향상시키는데 이용될 수 있다 (Turner and Brackbill, 1998) "


"웹 문서 작성자들은 검색과 통제를 향상시키기 위해서 keywords와 description 메타태그를 이용하는 것이 좋다"(AltaVista Search Network, 1997)
KISTI 한국과학기술정보연구원


Research Questions
웹 문서들은 어떠한 메타태그들을 포함하고 있을까?
웹상에서 주로 이용되는 메타태그들은 어떠한 것들일까?
메타태그들은 어떠한 형식으로 이용되고 있을까?
Keywords 및 Description 메타태그와 검색효율과는 어떠한 관계가 있을까?
KISTI 한국과학기술정보연구원

선행 연구

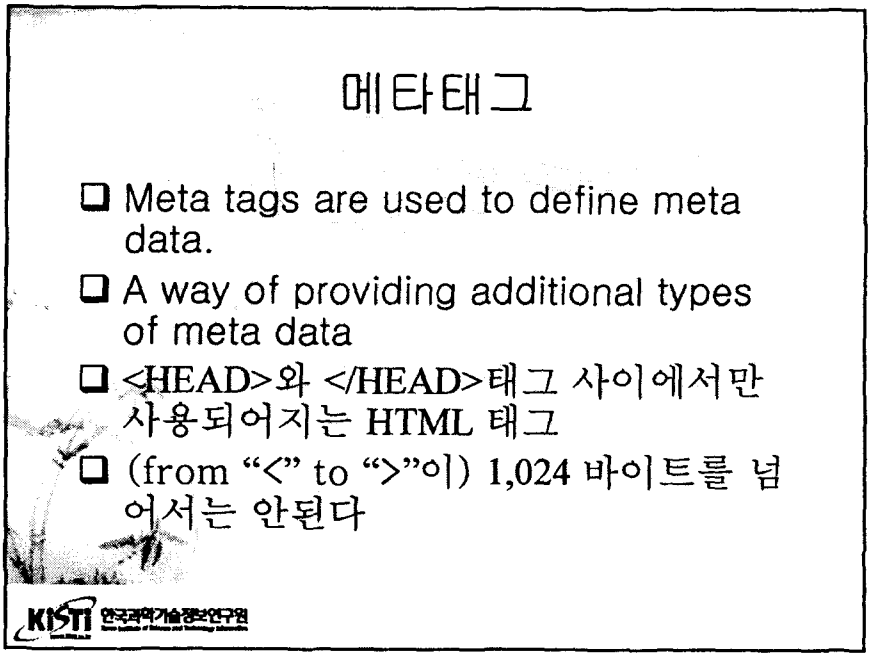
- Turner and Brackbill(1998)의 검색효율
  - Keywords 메타태그의 단독사용(O)
  - Description 메타태그의 단독사용(X)
  - Keywords 와 Description 메타태그의 공동사용(O)
- Danny Sullivan(1997)의 조사
  - Keywords 메타태그의 이용: 12%
  - Description 메타태그의 이용: 11%
- Scott Clark(1998)의 조사
  - Keywords and Description 메타태그의 이용: 21%

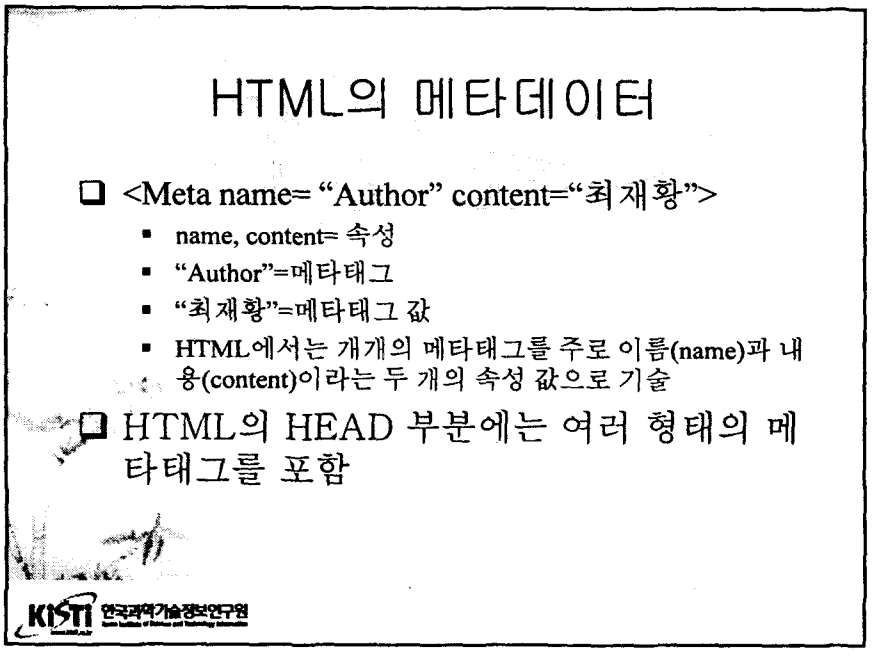
KiSTi 한국과학기술정보연구원

메타데이터

- Data about data
  - 실제 컨텐츠는 아니면서 이에 대한 각종 정보를 갖고 있는 데이터
- Its strength is not description but the support it provides for resource discovery(Lynch, 1998)
- Metadata prevents ambiguity about data(Lide, 1995)
- The centerpiece of information gathering(Weibel, 1995)
- machine understandable information for the web(W3C, 2000)

KiSTi 한국과학기술정보연구원


메타태그
Meta tags are used to define meta data.
A way of providing additional types of meta data
<HEAD>와 </HEAD>태그 사이에서만 사용되어지는 HTML 태그
(from "<" to ">"이) 1,024 바이트를 넘어서는 안된다
KISTI 한국과학기술정보연구원


HTML의 메타데이터
<Meta name= "Author" content="최재황">
name, content= 속성
"Author"=메타태그
"최재황"=메타태그 값
HTML에서는 개개의 메타태그를 주로 이름(name)과 내용(content)이라는 두 개의 속성 값으로 기술
HTML의 HEAD 부분에는 여러 형태의 메타태그를 포함
KISTI 한국과학기술정보연구원

HTML 메타데이터의 속성

- NAME 속성
- HTTP-EQUIV 속성
- CONTENT 속성
- SCHEME 속성

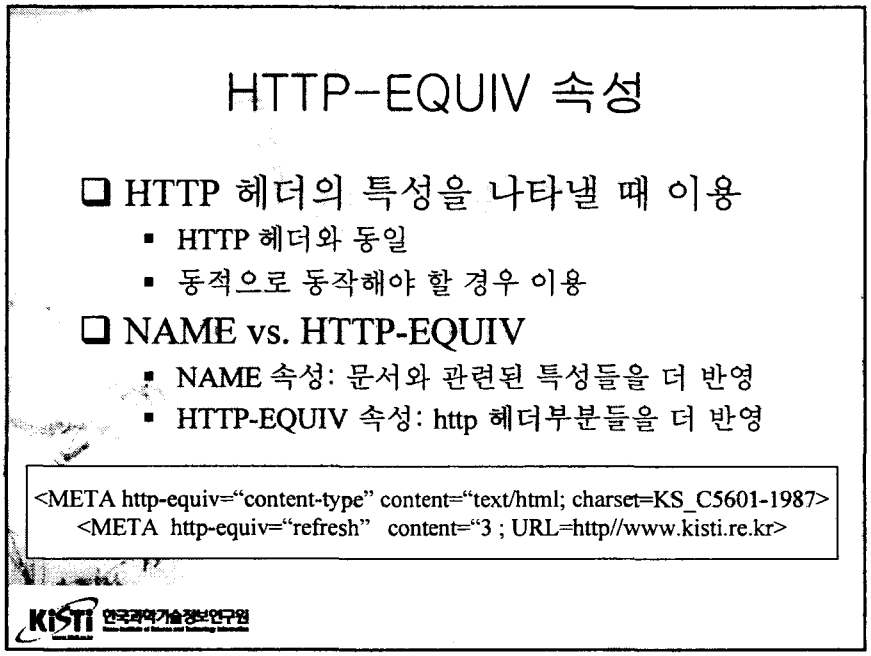
KISTI 한국과학기술정보연구원

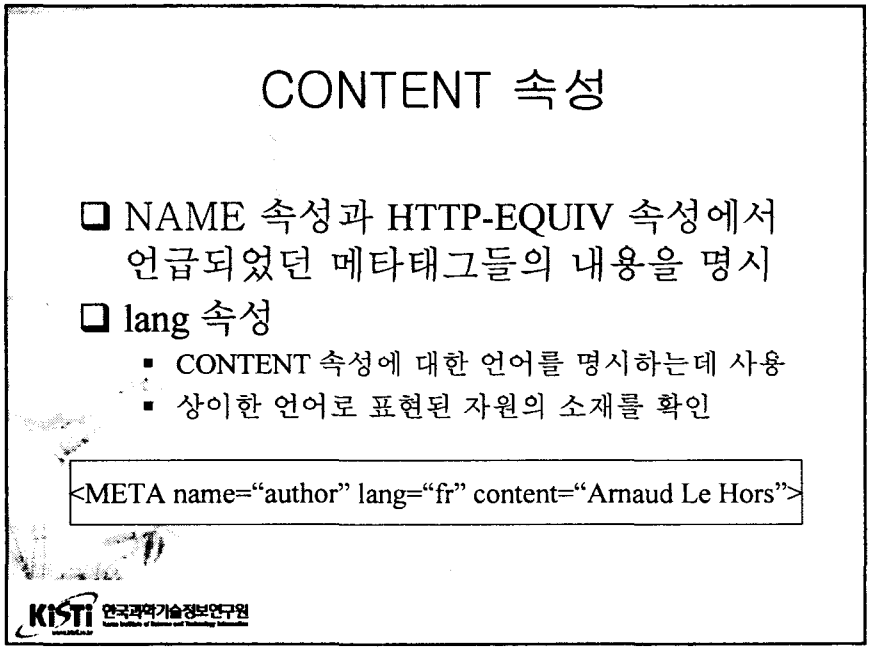
NAME 속성

- 메타태그의 이름을 표시
- HTML 명세는 메타태그들의 목록을 정의하지 않음
- 주로 Keywords와 Description 메타태그만 인식하고 이들을 색인에 이용

```
<META name="resource-type" content="document">
```

KISTI 한국과학기술정보연구원


HTTP-EQUIV 속성
HTTP 헤더의 특성을 나타낼 때 이용
HTTP 헤더와 동일
동적으로 동작해야 할 경우 이용
NAME vs. HTTP-EQUIV
NAME 속성: 문서와 관련된 특성들을 더 반영
HTTP-EQUIV 속성: http 헤더부분들을 더 반영
<META http-equiv="content-type" content="text/html; charset=KS_C5601-1987>
<META http-equiv="refresh" content="3 ; URL=http//www.kisti.re.kr>
KiSTi 한국과학기술정보연구원


CONTENT 속성
NAME 속성과 HTTP-EQUIV 속성에서 언급되었던 메타태그들의 내용을 명시
lang 속성
CONTENT 속성에 대한 언어를 명시하는데 사용
상이한 언어로 표현된 자원의 소재를 확인
<META name="author" lang="fr" content="Arnaud Le Hors">
KiSTi 한국과학기술정보연구원

SCHEME 속성

- CONTENT 속성에 나타난 값을 해석하거나 처리하기 위한 도구 제시
  - CONTENT에 수록된 값을 해석
  - 레코드의 일관성 유지
  - 표준화 달성
- 10-9-99는 무엇을 의미하는가?
  - 1999년 10월 9일?
  - 1999년 9월 10일?
  - Month-Day-Year

```
<META name="identifier" scheme="ISBN" content="0-8230-2355-9">
```

KiSTi 한국과학기술정보연구원

Description 메타태그(1)

- 해당 웹 문서의 내용을 요약
- 일부 검색엔진은 검색결과의 출력을 위해 Description 메타태그의 내용(content 부분)을 출력
- 일부 검색엔진들은 나름대로의 Description 메타태그에 대한 기준을 가지고 있다
  - 표준으로 정해진 것은 없다.

KiSTi 한국과학기술정보연구원

Description 메타태그(2)

- AltaVista, Hotbot, InfoSeek, Excite, WebCrawler 등
- 150 바이트에서 395 바이트까지 검색엔진에 따라 다양
  - InfoSeek: a limit of 200 characters
  - Hotbot: a limit of 150 characters
- Many search engines use only the first 200 characters.

KISTI 한국과학기술정보연구원

Keywords 메타태그

- 해당 웹 문서 내용의 키워드
- 키워드의 내용은 순위에 영향
- AltaVista, Hotbot, InfoSeek, WebCrawler 등
  - 정해진 표준은 없다.
  - Many search engines recommends that the keywords tag be between 300 and 500 characters.
- Excite
  - Description 메타태그 지원함
  - Keywords 메타태그 지원 안함

KISTI 한국과학기술정보연구원

Keywords와 Description 메타태그의 예

```
<HTML>
<HEAD>
  <TITLE>The Web Developer's Virtual Library</TITLE>
Keywords
  <META   NAME = "Keywords"  CONTENT="HTML, CGI, Java,
          VRML, browsers, plugins, graphics, HTTP servers, JavaScript, Perl,
          ActiveX, Shockwave">
Description
  <META   NAME="Description"  CONTENT=" Locate web
          authoring and software internet resources at The WDVL,
          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>
```

KiSTi 한국과학기술정보연구원

연구 데이터 추출(1)

- OCLC FirstSearch, NetFirst
- FirstSearch
  - 60여 개의 유형별, 분야별 데이터베이스
  - 150만 건 이상의 Full text 논문 수록
  - 3,600만 종의 도서 및 비도서 자료 검색
- NetFirst
  - 인터넷 상에서 이용할 수 있는 자원에 대한 신뢰성있고 권위있는 정보를 제공

KiSTi 한국과학기술정보연구원

연구 데이터 추출(2)

- Library, information, science의 세 키워드 이용
- 1,000/3,913 (2000년 12월 14일)
- NetFirst의 선정이유
  - 질적인 면에서 일반 검색엔진보다 우수
  - 메타태그를 더 많이, 정확하게 이용

KISTI 한국과학기술정보연구원

메타태그의 분포(1)

- 1회만 발생한 메타태그의 수: 33개
- 2회 이상 발행한 메타태그의 수: 59개
- 총 발생 메타태그의 수: 92(33+59)개

KiSTi 한국과학기술정보연구원

메타태그의 분포(2)

- 200회 이상 발생한 메타태그
  - Content-type: 250회(25.0%)
  - Generator: 229회(22.9%)
  - Keywords: 229회(22.9%)
  - Description: 206회(20.6%)
- Description or Keywords: 244개(24.4%)
- Description and Keywords: 191개(19.1%)

KiSTi 한국과학기술정보연구원

메타태그의 분석

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

KiSTi 한국과학기술정보연구원

Keywords 메타태그의 분석(1)

- 전체평균 메타태그 사이즈: 291 바이트
- 실제평균 메타태그 사이즈: 255 바이트
  - 0-300 바이트: 177/229(77.3%)
  - 300-500 바이트: 29/229(12.7%)
  - :
  - :
  - 1024+ 바이트: 8/229(3.5%)

KiSTi 한국과학기술정보연구원

Keywords 메타태그의 분석(2)

- 메타태그 이름의 다양성
  - keyword,keywords, KeyWords 등
  - DC.Subject(76회 발생), key phrases(2회 발생) 제외
- 메타태그의 잘못 사용
  - ">"를 빠뜨린 경우 등
- 메타태그의 중복 사용
- NAME 속성과 CONTENT 속성이 바뀐 경우
- CONTENT 속성에 내용이 없는 경우
- LANG 속성이 이용된 경우

KiSTi 한국과학기술정보연구원

Description 메타태그의 분석(1)

- 전체평균 메타태그 사이즈: 202 바이트
- 실제평균 메타태그 사이즈: 165 바이트
  - 0-200 바이트: 155/206(75.2%)
  - 201-500 바이트: 40/206(19.4%)

:

:

  - 1024+ 바이트: 1/206(0.5%)

KiSTi 한국과학기술정보연구원

Description 메타태그의 분석(2)

- 메타태그 이름의 다양성
  - description, Description, DESCRIPTION 등
  - DC.Description(69회 발생) 제외
- 메타태그의 잘못 사용
  - ">"를 빠뜨린 경우 등
- 메타태그의 중복 사용
- NAME 속성과 CONTENT 속성이 뒤빠뀐 경우
- CONTENT 속성에 내용이 없는 경우
- LANG 속성이 이용된 경우

KISTI 한국과학기술정보연구원

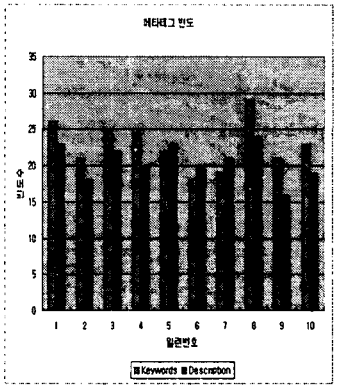
메타태그와 검색효율과의 관계(1)

- 먼저 검색된 웹 문서들은 나중에 검색된 웹 문서들 보다 더 많은 메타태그를 포함하고 있을까?
- 1;,000개의 웹 문서들을 10개(100개씩)로 나누고 각각의 그룹에 포함된 두 메타태그의 개수를 확인

KISTI 한국과학기술정보연구원

메타태그와 검색효율과의 관계(2)

일련번호	메타태그 빈도	
	Keywords	Description
1 - 100	26	23
101 - 200	21	18
201 - 300	25	22
301 - 400	25	20
401 - 500	22	23
501 - 600	18	20
601 - 700	19	21
701 - 800	26	24
801 - 900	21	18
901 - 1000	23	19
합 계	229	206



KiSTi 한국과학기술정보연구원

메타태그와 검색효율과의 관계(3)

일련번호	Keywords Tag Only	Description Tag Only	Keywords and Description Tags
0 - 100	4	1	22
101 - 200	6	3	15
201 - 300	5	2	20
301 - 400	5	0	20
401 - 500	2	3	20
501 - 600	1	3	17
601 - 700	0	2	19
701 - 800	5	0	24
801 - 900	6	1	15
901 - 1000	4	0	19
합 계	38	15	191

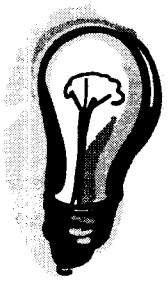
KiSTi 한국과학기술정보연구원

향후 연구과제

- 웹 DB별 메타태그 이용에 관한 비교연구
  - 메타태그를 인식하는 검색엔진
  - 일반(메타태그를 인식하지 못하는) 검색엔진
  - 상용 웹 DB

KISTI 한국과학기술정보연구원

감사합니다.



KISTI 한국과학기술정보연구원