

정보검색과 웹 3강

하 석 재

CEO, 2HCUBE

sjha72@gmail.com

지난 시간엔

- 웹기반 시스템의 구성
 - 1-tier/2-tier/3-tier/N-tier
 - WAS서버(로드밸런싱, 로컬/전역 트랜잭션처리, 고가용성, ...)
 - BEA Systems(Oracle) Tuxedo(미들웨어)/Weblogic(WAS)
 - Tmax JEUS(WAS)
 - 캐시서버(Proxy/Reverse Proxy)
 - 로드밸런서(L4/L7 스위치)
 - 스택 / 다이나믹
 - 방화벽(포트포워딩) / 웹방화벽
 - IDS/IPS
- 큐잉(카프카) / 고가용성(HA)

지난 내용 보완

- <https://sweet-rain-kim.tistory.com/220>
- <https://devocean.sk.com/experts/techBoardDetail.do?ID=165711>
- <https://devocean.sk.com/blog/techBoardDetail.do?ID=165737>
- Raft(합의)알고리즘
 - 일종의 복제(Replication) - Leader/Follower/Candidate
 - <https://always-kimkim.tistory.com/entry/digging-kraft-raft-algorithm>
- 카프카에서 HA필수관련 언급과련 추가 내용
- 주키퍼(Zookeeper) -> KRaft(크래프트)로 변경(2024, Kafka 4.0부터 기본)
 - 자체 알고리즘으로 교체 및 성능개선
 - 컨트롤러는 메모리에 메타데이터 캐시유, 메타데이터 동기화/관리과정 효율화
 - 메타데이터 복제시간 감소 -> 효율적인 컨트롤러의 리더 선출작업

수업일정

Date	Contents	
1주 (9/1)	오리엔테이션 , 과정소개	
2주(9/8)	웹기반 시스템의 구성(큐잉, 고가용성)	
3주(9/15)	HTML 1	
4주(9/22)	HTML 2	
5주(9/29)	HTML 3	
6주(10/6)	HTTP 1	온라인(추석)
7주(10/13)	HTTP 2	
8주(10/20)	중간고사	
9주(10/27)	검색엔진과 전문검색	
10주(11/3)	토큰라이저/임베딩벡터기반 검색	온라인
11주(11/10)	Lucene(TF/IDF)	
12주(11/17)	엘라스틱서치 1	
13주(11/24)	엘라스틱서치 2	
14주(12/1)	엘라스틱서치 3	
15주(12/8)	기말고사	

HTML 1



HTML

- WWW = HTML + HTTP
- 웹 표준 만드는 기관 - **w3c.org**
- 스펙 및 실습
 - <https://www.w3schools.com/>
- **HTML5(넓은 의미) = HTML5(좁은 의미) + CSS3 + ECMAScript(ES) + SVG(2D) + WebGL(3D)**

HTML

- 초기의 웹
 - 팀 버너스 리(CERN)
- 웹브라우저의 계보
 - NCSA Mosaic - Netscape - Firefox - **Chromium**
 - Internet Explorer - Edge - Edge(Chromium)
- Chromium
 - 오픈소스
 - Chrome(구글) / Edge(MS) / Whale(네이버) / 삼성브라우저 / ...

Chromium vs. Chrome

- 오픈소스 중 리눅스와 더불어 Big2
- 오픈소스 vs. 구글 브라우저
 - 구글, MS, 네이버, ... 등이 공동 기여후 성과공유
 - 방대한 코드 검증 및 개발비용 문제의 해결책
 - MS조자 자체 Edge개발을 포기하고 크로미움기반 Edge로 선회
- 개별 브라우저는 크로미움기반이지만 100% 동일하지 않음
 - cf. 구글 Android Open Source Project(AOSP) vs. 삼성 안드로이드
 - 버그해결 및 별도의 기능 추가
 - 특히 동영상 코덱부분 해결(av1, VP9, h.264/h.265/h.266, ...)

HTML

- 웹서버
 - Apache2 / Nginx
 - 스태틱 콘텐츠 실행
 - 주로 80번 포트를 사용
 - 암호화(HTTPS, TLS/SSL)를 사용할 때는 443을 사용
- 서블릿/JSP 엔진
 - Apache Tomcat / Resin
 - 다이나믹 콘텐츠 실행
 - 주로 8080번 포트를 사용
 - 암호화(HTTPS, TLS/SSL)를 사용할 때는 8443을 사용

웹 서버(스태틱 + 다이내믹) / 연동

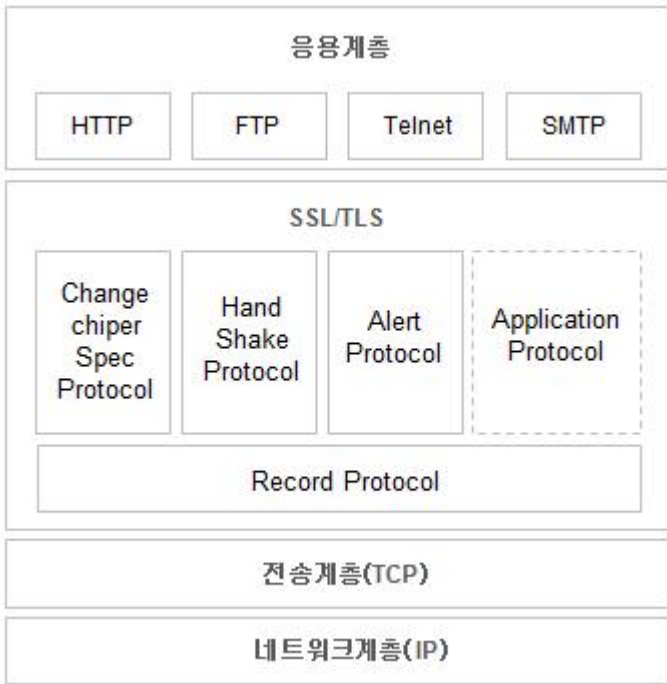
- HTML
 - <http://a.b.c.d:80/sample.html>
 - <https://a.b.c.d/sample.html>
- JSP/Servlet
 - <http://a.b.c.d:8080/sample.jsp>
 - <https://a.b.c.d:8443/sample.jsp>
- Apache + 톰캣 연동
 - <http://a.b.c.d:80/sample.jsp>

HTTP 암호화

- 7계층 암호화(ALS:Application Layer Security)
 - Secure HTTP
 - shttp://
 - 응용프로그램별 암호화
- 4계층 암호화(TLS:Transport Layer Security)
 - HTTP with SSL/TLS(443/8443)
 - https://
 - 포트별 암호화
- 3계층 암호화(NLS:Network Layer Security)
 - IP Security(IPsec)
 - IP별 암호화

SSL/TLS

- <https://m.blog.naver.com/xcripts/70122755291>
- Secure Socket Layer
 - 전송단 암호화 기술
- SSL/1.0 -> SSL/2.0 -> SSL/3.0(TLS/1.0으로 명칭변경)
- TLS/1.0 -> TLS/1.1 -> TLS/1.2 -> TLS/1.3
- 과거에는 선택적인 암호화(포트별)이었지만 최근에는 100% 적용
 - 시스템 부하 vs. 보안
 - HTTP/2, HTTP/3에서 필수가 됨(차후에 자세히 언급예정)



HTML

- <https://d2.naver.com/helloworld/59361>
 - 웹브라우저는 어떻게 동작하는가? cf. DOM/CSSOM, Webkit/v8 engine
- 웹 브라우저는
 - HTML 파서(해석기) + HTML 렌더러(페이지 그리기)
 - 전달된 HTML의 내용을 이해(파싱)하고 / 페이지를 실제로 만들어 출력(렌더링)
 - HTTP 클라이언트
 - 통신 클라이언트 담당
- 크로미움의 blink엔진
 - [https://ko.wikipedia.org/wiki/%EB%B8%94%EB%A7%81%ED%81%AC_\(%EB%B8%8C%EB%9D%BC%EC%9A%B0%EC%A0%80_%EC%97%94%EC%A7%84\)](https://ko.wikipedia.org/wiki/%EB%B8%94%EB%A7%81%ED%81%AC_(%EB%B8%8C%EB%9D%BC%EC%9A%B0%EC%A0%80_%EC%97%94%EC%A7%84))
 - Webkit의 웹코어 컴포넌트에서 fork(파생)됨

HTML 의 역사(+파편화)

- HTML 1 -> 2 -> 3(.2) -> 4(DHTML/Flex) -> 5
- HTML 4
 - CSS의 분화(내용과 스타일을 분리)
 - HTML3.2 / CSS 1.0 -> HTML4 /CSS 2 -> HTML5 / CSS3
 - 파편화(**Dynamic HTML**)
 - MS vs. Netscape(Firefox)
- UX(User Exprience)
 - RIA(Rich Internet Application) - 2000년 대 초반
 - Adobe Flex(flash의 변형)의 액션스크립트
- MS IE가 아닌 웹 표준 준수로 변경
 - 브라우저와 관계없이 실행가능하도록 -> 크롬으로 대세가 바뀜

HTML/CSS history

- https://www.w3schools.com/html/html_intro.asp

HTML

- Hypertext Markup Language
- Tag기반 언어
 - `<h1> ~ </h1>` -> **element(엘리먼트)**
 - `
` / `<hr/>`
- Well-formed vs Ill-formed
 - `
` vs `
` / 소문자 사용여부
 - XHTML vs. HTML
 - **Required, Mandatory, Compulsory(필수) / Recommended(추천/권고) / Optional**

실습관련

- 실습여부를 확인할 수 있도록 이미지 캡처(주소창 URL이 보여야 함)
 - 2-3주에 모아서 한꺼번에 제출(별도 안내)

HTML 파일(샘플)

<!DOCTYPE html>

<html>

 <head>

 <title>Hello World</title>

 </head>

 <body>

 <h1>Hello World</h1>

 </body>

</html>

- DOM트리 - https://www.w3schools.com/js/pic_htmltree.gif

HTML 파일(샘플)

```
<h1>Hello World</h1>
```

Doctype

- 문서의 스키마(또는 DTD)를 정의
- HTML5
 - **<!DOCTYPE html>**
- HTML 4
 - <http://www.tcpschool.com/html-tags/doctype>
 - `<!DOCTYPE HTML PUBLIC "-//W3C//DTD HTML 4.01//EN" "http://www.w3.org/TR/html4/strict.dtd">`
 - `<!DOCTYPE HTML PUBLIC "-//W3C//DTD HTML 4.01 Transitional//EN" "http://www.w3.org/TR/html4/loose.dtd">`
 - `<!DOCTYPE HTML PUBLIC "-//W3C//DTD HTML 4.01 Frameset//EN" "http://www.w3.org/TR/html4/frameset.dtd">`

Tags

- `<h1>` ~ `<h6>`
 - `<p>`
 - `<a>`
 - ``
-
- 태그명은 소문자로 **(recommended)**
 - XHTML(required)

Tags / Attributes

- `<h1> ~ <h6>`
- `<p>`
- ``
 - href
- ``
 - src
- 태그명/속성명은 소문자로(recommended)
 - XHTML(required)
- 속성명은 `key="value"` 형태로 cf. `checked="checked"`

HTML comments

- HTML내에 `<!-- -->` 넣으면 아무런 동작 안 함
`<!-- Write your comments here -->`

HTML Table

```
<table border="1">
  <tr>
    <th>name</th>
    <th>email</th>
  </tr>
  <tr>
    <td>seokjae ha</td>
    <td>sjha72@gmail.com</td>
  </tr>
</table>
```

- <tr>~</tr>, <td>~</td>, <th>~</th>
- colspan vs. rowspan

HTML List

- Ordered List, Unordered List

```
<ul>
```

```
  <li>seokjae ha</li>
```

```
  <li>sjha72@gmail.com</li>
```

```
</ul>
```

```
<ol>
```

```
  <li>seokjae ha</li>
```

```
  <li>sjha72@gmail.com</li>
```

```
</ol>
```

질문사항

- 수업관련 질문이 있으면
 - sjha72@gmail.com로 문의해 주세요

감사합니다

