

사용자 행동 패턴과 웹 문서 구조 분석을 이용한 웹 페이지 테스트 방법

유대승^o 심민석 엄정섭 이명재
울산대학교 컴퓨터·정보통신공학부
{yds, wolf, quaza, ymj}@cic.ulsan.ac.kr

A Method for Testing Web pages Using the Analysis of User Behavioral Pattern and Web Document Structures

Daesung Yoo^o Minsuck Sim Jeongseob Eom Myeongjae Yi
School of Computer Engineering & Information Technology, University of Ulsan

요 약

웹 기반 시스템의 테스트에 대한 정확성과 신뢰성을 향상시키는 것은 시스템의 질을 향상시킬 수 있고 시스템 개발 및 유지보수에 대한 비용을 현저히 줄일 수 있다. 본 논문에서는 웹 로그 분석을 통하여 테스트 환경 설정에 필요한 정보와 사용자들에 대한 패턴 정보를 추출하였다. 그리고 웹 콘텐츠 분석을 통하여 링크 정보와 데이터베이스에 대한 접근 또는 사용자 입력을 요구하는 모듈에 대한 정보를 추출하였다. 추출한 정보를 이용하여 테스트를 위한 시나리오 문서인 테스트 스크립트를 생성함으로써 실제에 근접한 테스트 시뮬레이션을 구현하는 방법을 제시하였다.

1. 서 론

웹의 빠른 발전과 더불어 비즈니스 프로세스, 지식 경영, 마케팅, 전자 상거래 등 현대의 다양한 IT 시스템들이 웹으로 통합되어 가고 있다. 웹 통합 개발 과정에 있어 필연적으로 파생되어진 복잡한 개발 및 유지보수를 더 어렵게 한다. 이에 따른 비용 또한 증가하고 있다[1, 2].

웹 기반 시스템에 대한 효율적인 개발과 유지보수 및 확장을 위해서는 시스템 개발의 생명주기에 따라 적절한 테스트 방법에 대한 연구가 필수적이다. 개발 단계에서는 개발중인 시스템에 존재할 수 있는 병목 현상과 데드락과 같은 문제점을 미리 예측함으로써 즉각적인 문제 해결이 필요하며, 테스트 단계에서는 가상 사용자 시뮬레이션을 통하여 실제 시스템 운영시에 발생할 가능성이 있는 문제점들을 시험할 수 있어야 한다. 또한 시스템을 배포하여 운영하는 단계에서는 운영중인 시스템에 대한 모니터링을 통하여 발생 가능한 문제점을 예측하고 미리 대처하거나 사용자 수의 증가에 따른 시스템 확장 계획 수립에 적용할 수 있어야 한다. 이와 같이 웹 어플리케이션의 테스트에 대한 연구는 시스템의 생명주기에 따라 발생 가능한 문제에 대한 조기 발견과 시스템의 성능 향상 및 시스템의 확장 계획 수립에 있어 중요한 선결 과제로 제기 되고 있다.

웹 테스트는 가상의 환경을 설정하고, 행동 정보를 가진 가상의 사용자들을 생성하여 실제 운영 환경과 유사한 시뮬레이션을 구현함으로써 이루어진다. 이런 가상의 환경과 사용자에 대한 정보를 저장하는 문서를 테스트 스크립트라고 한다. 웹 테스트를 위한 테스트 스크립트를 작성하는 방법으로 웹 브라우저상의 사용자 행동(마우스 클릭 또는 키보드 입력)을 레코딩하는 방법, 로그파일을 분석하는 방법, 디렉토리 구조 정보를 이용하는 방법 등이 제시되고 있지만 실제 운영 환경에 근접한 테스트 스크립트를 작성 하기는 어렵다. 특히, 사용자 입력을 필요로 하는 동적 페이지의 경우에는 매우 제한적이다.

본 논문에서는 정확하고 신뢰성 있는 웹 테스트를 위해서 웹 로그를 분석하여 테스트 환경 설정에 필요한 정보와 사용자 행동에 대한 패턴 정보를 추출하였다. 또한 웹 콘텐츠를 분석하여 링크 정보와 사용자 입력 요구에 대한 정보를 추출하였다. 그리고 추출한 정보를 이용하여 테스트 환경 설정과 가상 사용자 생성에 이용함으로써 웹 테스트를 위한 보다 향상된 방법을 제시하였다.

본 논문의 구성은 다음과 같다. 2장에서는 웹 테스트에 대한 설명과 웹 테스트에 대한 연구 및 테스트 도구들을 살펴본다. 3장에서는 테스트 스크립트의 생성 과정을 보인다. 4장에서는 테스트 스크립트 생성 과정을 단계별로 세부적으로 설명한다. 끝으로 5장에서는 결론 및 향후 연구방향에 대해 살펴본다.

2. 웹 테스트 및 웹 테스트 도구

2.1 웹 테스트

웹 테스트는 웹 기반 시스템에 대한 테스트로 주로 성능 테스트를 의미한다. 성능 테스트는 로드 테스트, 스트레스 테스트, 내구성 테스트 등으로 분류할 수 있다[3]. 그 외 사용자 환경 테스트, 웹 브라우저 렌더링 테스트, 링크 및 구문 검사, 사용성 및 신뢰성 테스트 등이 있다 [4].

웹 기반 시스템은 기존의 시스템과는 다른 특징들(많은 사용자수, 다양한 사용자 환경, 통합된 시스템, 문서 기반)을 가지므로 기존의 테스트 방법과 도구들을 그대로 이용할 수 없다. 따라서 웹 테스트에 대한 많은 새로운 연구가 진행중이다. 그 중에서 전체 웹사이트를 테스트 가능한 컴포넌트(또는 객체)로 분리하고 이를 반복적으로 테스트함으로써 전체 웹사이트를 테스트하는 방법[3]과 접근로그(access log)를 이용하여 통계학적 테스트를 위한 모델을 만들고 에러로그(error log)를 분석하여 신뢰성을 검증하는 연구[4] 등이 있다.

본 연구는 정보통신부의 "정보통신 우수시범학교 지원사업"의 지원에 의해 이루어졌습니다.

2.2 웹 테스트 도구

웹 테스트를 위한 도구들 중에서 다른 것들과 비교되는 특징을 가진 것으로는 Rational사의 SiteLoad[5], Envice사의 Prophecy[6], 그리고 Microsoft사의 WAS[7]가 있다.

Rational사의 SiteLoad는 웹 브라우저 기반의 사용자 인터페이스를 제공함으로써 사용하기 쉬우며, 다양한 플랫폼을 지원한다. 또한 자체 웹 브라우저를 이용하여 사용자 액션을 기록하고 기록된 정보를 이용하여 테스트 스크립트를 작성하고 있다.

Envice사의 Prophecy는 테스트 스크립트 작성을 위한 클라이언트 프로그램을 로컬 컴퓨터에 설치하고 작성된 스크립트를 자사의 서버에 업로드한 후 테스트를 수행한다. 이것은 일반적으로 로컬 호스트에서 테스트를 실행하는 방법과는 비교되는 특징이다.

Microsoft사의 WAS는 사용자 액션 레코딩, 로그파일 분석, 링크 분석 등을 통하여 다양한 방법으로 테스트 스크립트를 작성할 수 있다. 그리고, 테스트를 위한 여러 가지 성능 카운터를 임의적으로 세팅할 수 있어 테스트 목표에 따라 다양한 모니터링을 수행할 수 있다.

3. 테스트 스크립트 생성 과정

실제 시스템의 운영 상황과 유사한 시뮬레이션을 위하여 웹 로그를 분석하여 사용자 행동 패턴을 추출하고 웹 콘텐츠의 구조 분석을 통한 링크 정보 및 구조 정보 추출을 수행한다.

추출한 사용자 행동 패턴 정보와 테스트 환경 설정 값과의 연관 관계를 도출하고 링크 정보와 문서 구조 정보를 이용하여 정적 페이지 및 동적 페이지 테스트를 위한 테스트 스크립트를 생성한다. 테스트 스크립트에는 가상 사용자의 행동 정보와 작업 로드를 생성하는 로드 생성 클라이언트의 행동 정보 및 환경설정 정보를 표현한다.

그림 1은 본 연구에서 수행한 테스트 스크립트 생성 과정을 보여준다.

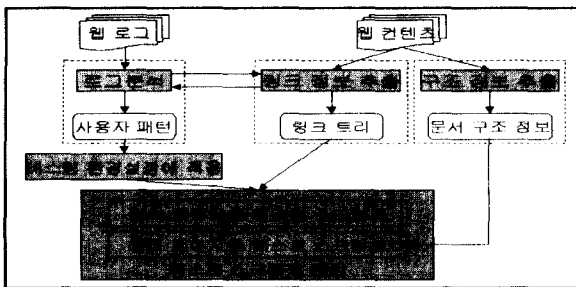


그림 1 테스트 스크립트 생성 과정
스트 스크립트 생성 과정을 정리하면 다음과 같다.

- 로그 분석을 통한 사용자 패턴 추출 : 패턴 추출을 위해서 로그에 기록된 정보(날짜, 시간, ip, 메서드, 접근 문서 정보, 서비스 상태값 등)들을 분리 저장하고 이 정보를 이용하여 사용자 행동에 대한 패턴 정보 추출
- 링크 정보 추출 : 웹 브라우저의 웹 문서 요청을 통한 링크 정보 추출
- 시스템의 구조 정보 추출 : 웹 콘텐츠 정보를 이용하여 시스템의 구조 정보(사용자 입력, 데이터베이스 접근 등)를 추출
- 테스트 스크립트 생성 : 추출된 사용자 행동 패턴, 링크 정보, 구조 정보를 이용하여 실제 상황에 근접한 정적, 동적 페이지에 대한 가상 사용자 스크립트 생성

4. 단계별 세부 내용

4.1 로그 분석

현재 대표적인 웹 서버에는 MS의 IIS와 Apache 서버가 있다. IIS는 W3C 확장 로그, ODBC 로그, NCSA 로그, IIS 로그파일 형식을 지원하고 Apache는 NCSA 로그, Apache 로그파일 형식을 지원한다. 본 연구에서는 두 웹서버에서 공통적으로 지원하는 NCSA 로그파일의 형식을 기준으로 연구를 수행 하였다.

NCSA 로그에 기록된 날짜, 시간, IP 주소, 메서드, 접근 문서 정보, 서비스 상태값 등의 정보들을 다음 단계인 패턴 분석의 기초 자료로 사용하기 위하여 그림 2와 같이 추출하고 저장한다.

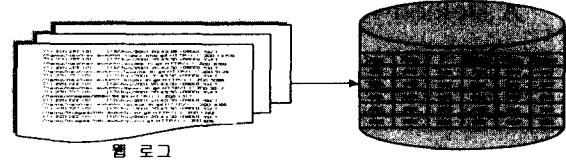


그림 2 로그 분석 과정

또한 로그 정보는 사용자 행동을 순차적으로 기록하므로 웹 페이지의 링크 정보를 내포한다. 따라서 그림 3과 같이 4.3절의 웹 콘텐츠로부터 링크 정보를 추출하는 과정에 이용하면 더 정확한 링크 정보를 추출할 수 있다. 그리고 4.3의 사용자 행동 패턴 추출 과정에서 웹 페이지 당 요청 수를 계산하여 로드가 집중되는 페이지를 식별할 수 있다.

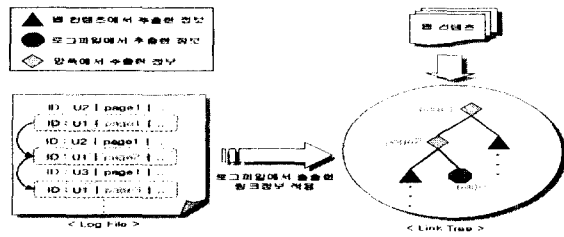


그림 3 로그파일 분석정보를 링크 정보 추출에 이용

로그 및 웹 콘텐츠로부터 추출한 링크 정보를 분류하면 다음과 같다.

표 1 추출한 링크 정보의 분류

웹 콘텐츠에서만 추출한 정보	링크 추출 오류이거나 요청이 없는 링크
로그에서만 추출한 정보	링크 추출 과정에서 빠진 링크이므로 링크 정보를 보정
양쪽 모두에서 추출한 정보	정확한 링크 정보

4.2 사용자 행동 패턴 추출

로그 분석을 통해 추출한 정보를 사용자별 패턴(문서 이동경로, 문서당 요청수, 문서당 머문시간 등)과 그룹별 패턴(동시접속자수, 문서당 중요요청수, 문서당 총 머문시간 등)으로 구분해서 패턴 정보를 추출한다. 추출한 정보는 테스트 환경 설정과 가상 사용자에 대한 행동 정보 생성에 중요하게 이용된다.

4.3 웹 콘텐츠의 정보 추출

웹 콘텐츠로부터 링크 정보와 문서 구조 정보를 추출한다. 정확한 링

크정보는 웹 브라우저의 문서 요청을 통하여 추출한 링크 정보와 4.1절에서 설명한 것과 같이 로그 분석 과정에서 추출한 링크 정보를 통합하여 얻을 수 있다. 링크 정보는 링크 트리를 형성하고 이는 사용자 패턴 정보와 함께 테스트 스크립트 생성에 이용된다. 링크 추출 과정에서는 추가적으로 끊어진 링크 및 사용되지 않는 문서 정보를 얻을 수 있다. 문서 구조 정보를 추출하는 목적은 동적 페이지 테스트에서 발생하는 문제점을 해결하기 위해서 그림 4와 같이 사용자 입력 폼, 값 전달 방식(POST, GET, HEAD), 데이터베이스에 접근 여부 등의 정보를 추출하는 것이다.

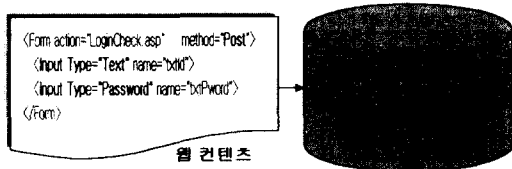


그림 4 사용자 입력 정보 추출

4.4 정적 페이지 및 동적 페이지 테스트 스크립트 작성

사용자 패턴 스크립트는 웹 테스트를 위해서 가상의 사용자가 가지는 행동 정보와 테스트를 위한 환경 설정에 대한 정보를 기술하는 문서로 웹 브라우저상의 사용자 액션(마우스 클릭 또는 키보드 입력)을 레코딩하는 방법, 로그파일을 분석하는 방법, 디렉토리 구조 정보를 이용하는 방법 등이 있지만 다양한 상황을 고려하지 않으며 동적 페이지를 테스트하기에는 부족한 점이 많다.

본 논문에서는 이와 같은 문제를 해결하기 위해서 통합적인 방법을 제시한다. 전 단계에서 추출한 문서 구조 정보, 로그 분석 정보, 링크 정보를 통합적으로 이용하여 그림 5와 같이 테스트 스크립트를 작성한다.

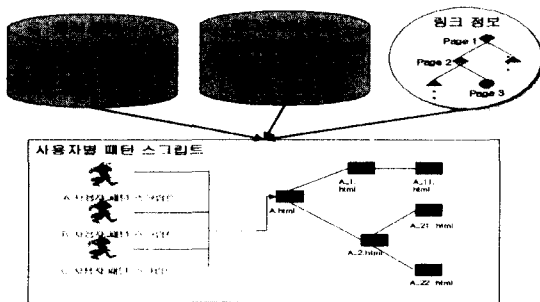


그림 5 사용자 패턴 스크립트 생성

로그 분석 정보와 링크 정보를 이용하여 정적 페이지에 대한 테스트 스크립트를 작성하고 문서 구조 정보를 고려하여 동적 페이지 테스트 시에 문제를 발생시키는 주요 요인인 사용자 입력 폼, 값 전달 방식(POST, GET, HEAD), 데이터베이스에 접근 등을 해결하는 동적 페이지에 대한 테스트 스크립트를 생성한다.

다음은 테스트 스크립트의 간단한 예이다.

```

<group name="default">
  <duration>
    <minute>10</minute>
  </duration>

```

```

<stresslevel>50</stresslevel>
<stressdelay>10</stressdelay>
<testlist>
  <testitem>
    <method>post</method>
    <path>http://203.250.77.151/login.asp</path>
    <thinktime>5</thinktime>
    <attlist>
      <attitem>
        <name>id</name>
        <value>guest</value>
        <name>pd</name>
        <value>guestpassword</value>
      </attitem>
    </attlist>
  </testitem>
  ...
</testlist>
</group>

```

5. 결론 및 향후 연구과제

웹 로그 분석을 통하여 테스트 환경 설정에 필요한 정보와 사용자들의 행동 패턴에 대한 정보를 추출하였다. 그리고 웹 콘텐츠 분석을 통하여 링크 정보와 데이터베이스 접근 또는 사용자 입력을 요구하는 모듈에 대한 정보를 추출하였다. 추출한 정보를 이용하여 테스트를 위한 가상 사용자를 생성하고 테스트 환경 설정에 이용함으로써 실제와 근접한 시뮬레이션을 구현하여 테스트의 정확성 및 신뢰성을 높일 수 있도록 하였다.

향후 연구과제로는 향상된 테스트 스크립트 문서 형식에 대한 연구가 수행되어야 한다. 테스트 스크립트는 가상 사용자에 대한 행동 정보와 가상의 로드를 생성시키는 로드 생성 클라이언트에 대한 행동 정보 및 환경 정보를 효과적으로 표현할 수 있어야 한다. 따라서 행동 정보를 표현하기 위해서 프로그래밍 언어적인 특성과 환경 정보를 표현하기 위하여 환경 설정 문서의 특성이 함께 갖추어져야 할 것이다.

[참고문헌]

- [1] Benoit Leger, Jean-Christophe Cimetiere, "Web Load and Performance Testing Tools", "www.trendmarkers.com", 2000
- [2] Hung Q. Nguyen, "Testing Applications on the Web", Wiley Computer Publishing, 2001
- [3] B.M. Subraya, S.V. Subrahmanya "Object driven Performance Testing of Web Applications", IEEE, 2001
- [4] Chaitanya Kallepalli, Jeff Tian, "Usage Measurement for Statistical Web Testing and Reliability Analysis", IEEE, 2001
- [5] Rational SiteLoad "www.rational.com/"
- [6] Envive Prophecy "www.envive.com/"
- [7] Microsoft WAS "webtool.rte.microsoft.com/"