논문 2011-2-2

추상화-여과-비교에 기반한 감정서 작성방법

전병태*

A Method for Making Appraisal Report Based on Abstraction-Filtering-Comparison

Byung Tae Chun*

요 약

유사도 감정은 감정 분야에서 많은 비중을 차지하고 있다. 따라서 효과적인 유사도 감정서를 작성을 위한 방법이 마련되어야한다. 본 논문에서는 Learned Hand 판사에 의하여 제안된 추상화(abstraction)-여과(filtration)-비교(comparison) 이론을 유사도 감정에 적용해보았다. 추상화 단계란 소스 프로그램을 어떻게 분석할 것인가에 대한 구체적인 분석 방법 및 기준 제시로 볼 수 있으며, 여과 단계란 감정시 배재하여야할 항목 및 내용을 말한다. 비교 단계란 여과된 프로그램을 도구를 이용하여 분석하는 것을 말한다.

Abstract

Similarity appraisal accounts for a great part of appraisal area. Therefore, we need effective method for similarity appraisal. The judge, Learned Hand, suggested a theory called abstraction - filtration - comparison, we apply the theory to similarity appraisal. The abstraction step are analysis methods and criteria for analyzing source program. The filtration step to exclude contents and items in source programs. The comparison step analyze filtered source programs using analysis tool.

한글키워드 : 추상화, 여과, 비교, 감정서 작성

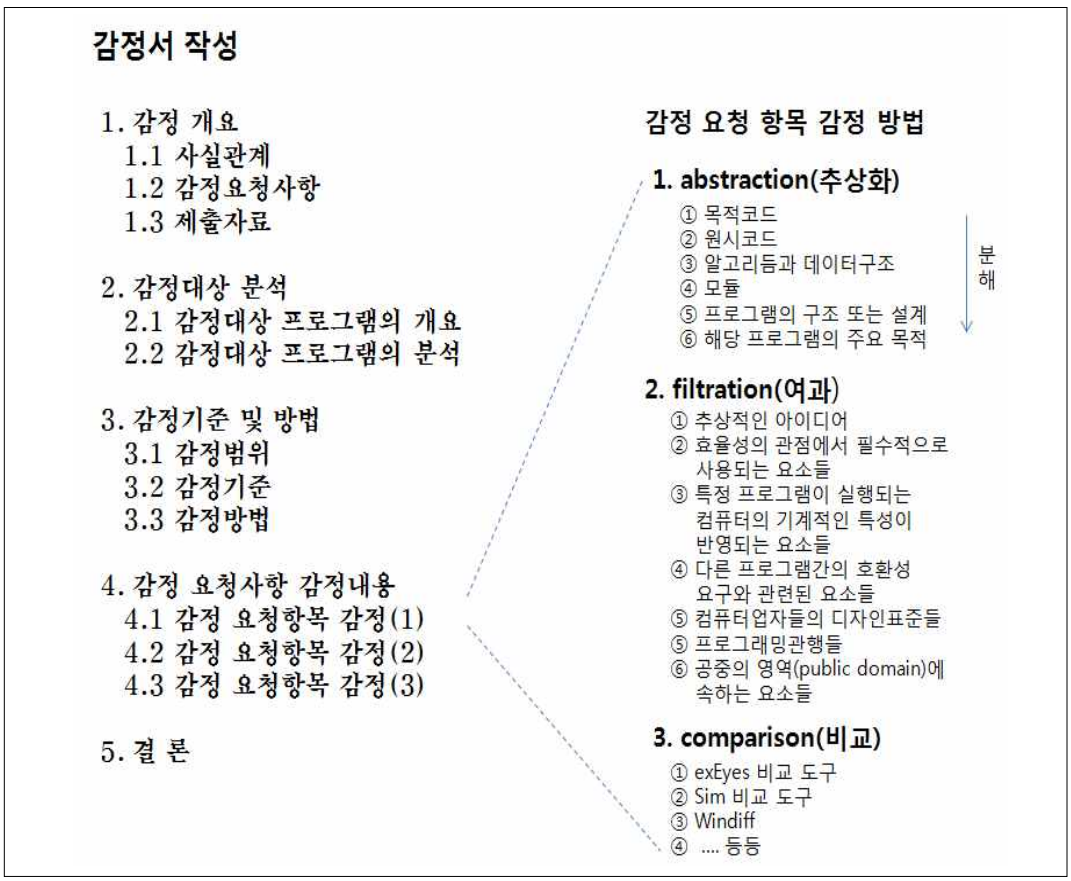
Ⅰ. 서 론

추상화 단계란(abstraction)는 일찍이 Learned Hand 판사에 의하여 주장된 추상화이론(abstraction test)을 컴퓨터 프로그램에 적용한 것으로, 프로그램의 작성과정을 문헌적 코드(literal code)에서부터 프로그램의 궁극적 기능(ultimate function)까지 작성과정의 역순으로 '추상화의 단계'를 분석하는 것을 말한다. Gates Rubber 사건의 판결에 의하면 추상화의 단계는 일반적으로 6단계로 나누어진다고 한다. 즉 ①목적코드, ②원시코드, ③알고리듬과 데이터구조,

* 한경대학교 웹정보공학과
(email: chunbt@hknu.ac.kr)
접수일자: 2011.10.5 수정완료: 2011.11.13

④모듈, ⑤프로그램의 구조 또는 설계, ⑥해당 프로그램의 주요 목적의 단계이다.

여과(filtration) 단계는 법원은 보호받을 수 없는 요소들을 하나씩 제거하여야 한다. 그것은 ① 추상적인 아이디어에 해당하는 요소를(아이디어와 표현의 합체이론 적용), ② 효율성의 고려에 의하여 지배되는 요소들(아이디어와 표현의 합체이론 수정·보완)이것은 이른바 합체(merger)이론을 컴퓨터 프로그램에 대한 관계에서 수정 보완하는 측면을 가지고 있다. 즉 어떠한 기능을 수행하기 위한 프로그램기술방법이 여러 가지가 있을 경우에도 그 방법들 사이에 효율성에 차이가 있어 결국 효율성이 높은 방법은 한 가지 뿐이라고 하게, 표현과 아이디어의 합체를 인정할 수 있게 된다. ③ 외적인 요인(external factors)에 의하여 지배되는 요소들이다. 위 ③의 요소들을 제거하여야 한다는 것을 Gates Rubber 사건 등의 일부 판례에서는 컴퓨터 프로그램보호에 있어서의 '표준적 삽화의 원칙'(scenes a fair doctrine) 어떠한 역사적 또는 가공적 주제를 다룸에 있어서는 특정의 표준화된 문학적 표현들이 도구로서 사용되지 않을 수 없는 경우가 있는데, 그러한 표현들을 저작권보호대상에서 제외하는 이론을 표준적 삽화의 원칙(scenes a fair doctrine)이라고 한다. 이것을 컴퓨터 프로그램보호에 있어서 적용할 때는 프로그래머의 선택가능성을 초월한 외부적 조건들을 비교대상에서 여과하는 것을 말할 때 이 용어를 사용하고 있는 것



[그림 1.1] 추상화-여과-비교에 기반한 감정서 작성방법

이다. 이에 해당하는 것으로 I)특정한 프로그램이 실행되는 컴퓨터의 기계적인 특성, ii) 어떤 프로그램이 상호 결합하여 사용될 것으로 예정하고 있는 다른 프로그램과 사이의 호환성 요구, iii)컴퓨터제조업자들의 디자인표준들, iv)그 프로그램이 제공되는 산업계의 요구, v)컴퓨터산업 내에서 널리 받아들여지는 프로그래밍 관행들을 들고 있다. 법원은 또한 公有(public domain)에 속하는 요소를 제외하여야 한다.

비교(comparison) 단계는 보호될 수 없는 요소들이 모두 걸러지고 나면 보호되는 표현의 핵이 남게 될 것인데, 이것을 피고의 작품과 비교하여야 한다고 한다. 보호되는 표현의 '핵'(core, kernel, nugget)이라는 용어를 사용함으로써 보호의 범위를 가능한 제한하고자 하는 의도를 보이고 있다.

II. 유사도 감정의 추상화(abstraction)

2.1 동종 언어간 유사도 감정의 추상화

동종 언어간 유사도 감정 추상화는 다음과 같다.

(1) 감정 도구인 exEYEs을 이용해 동일 또는 유사한 영역을 검출한다.

(2) 감정인이 직접 1단계에서 검출된 동일 또는 유사한 영역에 대해 2차적인 판단을 통해 확인한다.

(3) 감정 항목에 대한 평균 원본 또는 비교본 기준 유사도는 각 파일이 독립된 프로그램이므로 산술 평균한 값으로 산출한다. 또한 원본과 비교본 기준 유사도의 평균값도 두 값의 산술 평균으로 산출한다.

[표2.1] 는 각각의 비교 대상에 대해 유사도를 보여주고 있다.

[표 2.1] 유사도 측정예

순번	원본 파일명 (.ASM)	정제 라인수	비교본 파일명 (.ASM)	정제 라인수	동일 라인수	유사 라인수	원본기준 유사도 (%)	비교본 기준 유사도 (%)
1	DCTOAC91	561	Dctoac81-1	384	76	24	17.8	26.0
2			Dctoac81-2	412	70	26	17.1	23.3
3			change00	227	2	8	1.8	4.4
4	Dctoac94	615	Dctoac81-1	384	71	18	14.5	23.2
5			Dctoac81-2	412	65	20	13.8	20.6
6			change00	227	2	2	0	0
평균	갑제1-1호	588	을제2-1,2,3호	341	47.7	16.3	10.9	18.8

2.2 이종 언어 간 유사도 감정의 추상화

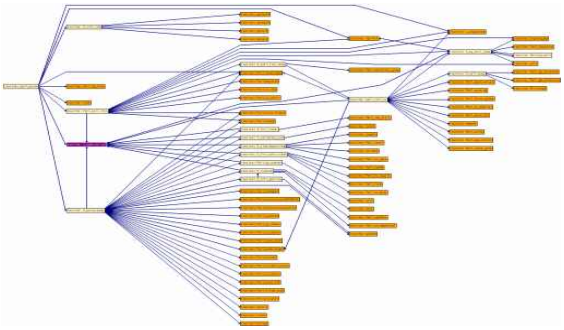
이종 언어간 유사도 비교는 소스 프로그램이 다르기 때문에 소스간 물리적 유사도 비교는 불가능하다고 볼 수 있다. 따라서 대부분 감정에서 소스 프로그램간 감정은 수행하지 않는다고 볼 수 있다. 이종 언어간의 유사도는 다음과 같은 요소에 대하여 감정을 수행한다.

- 파일 이름 및 디렉토리 구조
- Function Call Graph
- User Interface Design
- 소스 코드
- 주석 및 매뉴얼 및 유관 문서

제출된 프로그램이 실행 가능할 정도로 각 함수간의 관계가 잘 구성되었다면 함수간 호출(call graph) 관계를 이용하여 유사도 감정을 수행할 수도 있다. 그림 2.1은 Perl 언어에서 제공하는 함수 호출 도구를 보여주고 있다.

2.3 인터페이스 유사도 감정의 추상화

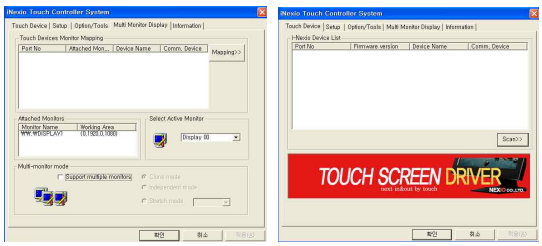
사용자 인터페이스 유사도 감정의 추상화는 요구 기능에 대한 인터페이스를 표[2.2]와같이 분류하고 그 기능 구현 및 동작 상태를 점검한다. 그림 2.2는 기능 중 멀티 모니터 설정 기능에 대한 점검을 나타낸다.



[그림 2.1] Call graph of the file perl.c

[표 2.2] 인터페이스 분류의 예

기능	A社	B社
멀티모니터 설정 기능	O	O
터치스크린 연결 상태 표시 기능	O	O
help 정보 표시 기능	O	O
그래픽 출력기능	O	O
멀티터치 제어 설정 기능	X	O
팝업 제어 기능	X	O



[그림 2.2] 멀티모니터 설정 기능

2.4 데이터베이스 유사도 감정의 추상화

데이터베이스에 대한 유사 여부는 원본(고소인 자료) 기준으로 원본 데이터베이스 전체의 테이블 중에서 피고소인 자료중 동일 명칭 테이블 개수의 비율을 산정한다.

테이블명 유사도(%) = 동일한 명칭 테이블 개수 / 원본 데이터베이스 전체 테이블 개수 x 100

테이블 명이 동일할 경우, 테이블 별 인덱스 유사도와 테이블 별 칼럼명 유사도는 피고소인 제출자료 및 송파구보건소의 테이블 인덱스나 칼럼 중, 원본 테이블의 인덱스나 칼럼명과 동일한 개수를 각 테이블 별 비율로 산정한다.

테이블 별 인덱스 유사도(%) = 테이블 별 동일 인덱스 개수/ 원본과 명칭이 동일한 각 테이블 별 인덱스 개수 x 100

테이블 별 칼럼명 유사도(%) = 테이블 별 동일 명칭 칼럼 / 원본과 명칭이 동일한 각 테이블 별 칼럼 수 x 100

데이터베이스 전체 인덱스(칼럼명) 유사도(%) = 테이블 별 인덱스(칼럼명) 유사도가 **'유사' 이상**인 테이블 수 / 전체 테이블 개수 X 100

테이블 별 칼럼명 유사도 판단 시, 데이터 길이 속성이 불일치하더라도 칼럼명이 동일하다면 동일 명칭 칼럼에 포함시킨다.

사용자 정의 프로시저는 데이터베이스의 자료를 참조하거나 수정할 때 흔히 사용되는 기능을 SQL(structured query language; 데이터베이스 질의 언어)로 미리 작성해둔 것을 뜻한다. 사용자 정의 프로시저도 데이터베이스 내에 저장되어 있는데, Microsoftn SQL Server 프로그램을 이용하면 데이터베이스 스키마와 사용자 정의 프로시저를 데이터베이스로부터 추출할 수 있다. 데이터베이스로부터 추출한 스키마와 사용자 정의 프로시저의 규모를 계산하면 표 [2.3]와 같다.

[표 2.3] 데이터베이스 유사도 감정 예

순번	감정대상 종류	원본 대상라인	비교본 대상라인	유사라인	원본기준 유사도	비교본기준 유사도
1	ASP와 JSP	85,890	53,513	10,674	12.43%	19.95%
2	데이터베이스 스키마	1,212	1,338	1,201	99.09%	89.76%
3	데이터베이스 사용자 정의 프로시저	612	6,976	611	99.84%	8.76%
종합 유사도		87,714	61,827	12,486	14.23%	20.20%

2.5 하드웨어 및 임베디드 소프트웨어 유사도 감정의 추상화

하드웨어 유사도 감정의 추상화는 표2.4와 같이 하드웨어를 기능별로 분류하여 추상화를 진행한다.

[표 2.4] 하드웨어 분류

대분류	소분류	기능	가중치
하드웨어부(30)	입출력부(10)	입력부(LCD touch, button)	0.5
		출력부(LCD, sound, vibration)	0.5
	메모리부(5)	processor model	0.2
		EEPROM	0.4
		외부 SDcard/메모리스틱	0.4
	통신부(5)	직렬통신(USB, RS232)	0.5
		무선통신(Bluetooth, WiFi)	0.5
	외장부(10)	외부디자인	0.5
		회로구성 architecture	0.5

[표 2.5] 소프트웨어 분류

대분류	소분류	기능	가중치
소프트웨어부(70)	응용프로그램(30)	정보기기 관리SW	0.5
		정보기기 응용SW	0.5
	디바이스 드라이버(30)	디바이스 설정	0.2
		I/O 프로그램	0.4
		통신프로그램	0.4
	펌웨어부(10)	유저인터페이스	0.2
		통신 프로그램	0.6
		입출력인터페이스	0.2

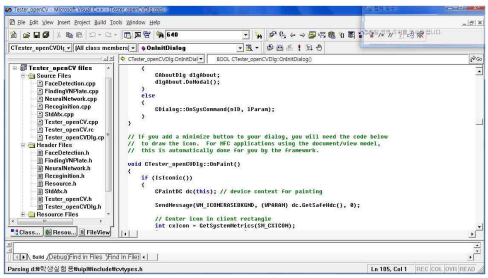
III. 유사도 감정의 여과(filtration)

3.1 효율성의 관점에서 필수적으로 사용되는 요소들

(1) 주석 제거

프로그램어가 프로그램 작성 시기, 프로그램 모듈 설명, 사용 방법 등을 기록하기 위하여 주석을 삽입한다. 이 부분은 프로그램으로 볼 수 없으므로 제거한다.

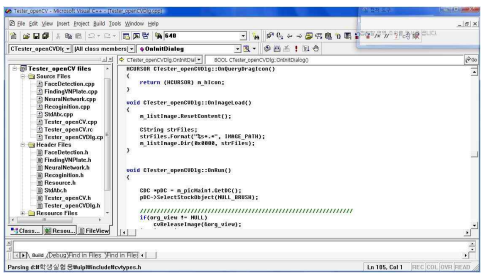
주석은 프로그램 유사도 산정시 제거하나, 주석은 정성적인 감정 평가에는 중요한 자료로 사용한다. 주석에는 사용자가 작성한 내용(프로그램 명, 작성 연월일, 작성자, 프로그램의 설명 등)은 그 프로그램의 독특한 자취가 나타날 수가 있다. 이러한 내용이 상대방 프로그램에 그대로 나타난다면 이는 프로그램을 복제/개정했다는 정황적 증거가 될 수 있다.



[그림 3.1] 주석 제거

(2) 공백 라인 제거

프로그램의 가독성을 높이기 위하여 삽입된 공백라인은 프로그램이라 볼 수 없으므로 공백라인을 제거하여야 한다.



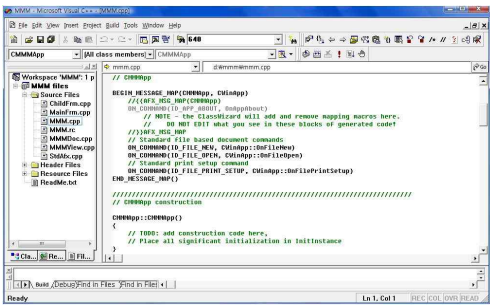
[그림 3.2] 공백라인 제거

3.2 특정 프로그램이 실행되는 컴퓨터의 기계적인 특성이 반영되는 요소들

(1) 자동 생성 코드 제거

비주얼 프로그램의 경우 클릭 몇 번으로 기본

프로그램이 작성된다. 이 프로그램은 프로그램에서 자동으로 만들어진 자동 생성 코드라고 볼 수 있다.



[그림 3.3] 자동 생성 프로그램

(2) 프로그램 특성상 사용되는 코드들

사용되는 함수나 관련 명령어들은 특정 프로그램이 실행되는 컴퓨터의 기계적인 특성이 반영되는 요소들은 제거한다. 특정 제조 업체가 제공한 소프트웨어는 제거한다.

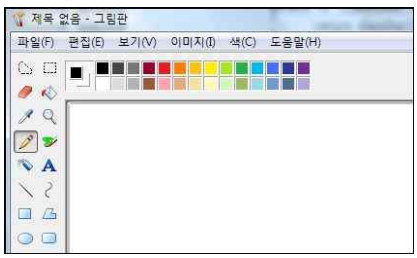
3.3 다른 프로그램간의 호환성 요구와 관련된 요소들

비주얼 베이직에서 c 함수를 호출할 경우 다른 프로그램간의 호환성과 관련된 명령어 또는 함수들이 사용된다. 호출 프로그램에서 프로그램머가 수정 가능한 것은 창작성이 있는 것으로 제거 대상에서 제외된다.

3.4 컴퓨터 업자들의 디자인 표준

인터페이스 디자인을 보면 업계에서 보편적으로 사용하는 메뉴 및 버튼이라 볼 수 있다. 이러한 공통 디자인 인터페이스는 공공적인 성격이 강하므로 창작성에서 이를 제외한다.

[그림 3.4] 동영상 플레이어 인터페이스



[그림 3.5] 그림판 메뉴 및 인터페이스

3.5 공중의 영역(public domain)에 속하는 요소들

① OpenSouce 제거

'오픈 소스 이니셔티브 (OSI:The Open Source Initiative)'라는 단체에 의해서 OSD(The Open Source Definition)라는 오픈소스에 대한 정의가 발표되었다. 이 정의는 '자유로운 재배포의 허가', '파생소프트웨어 배포의 허가', '개인이나 집단의 차별금지', '적용분야 제한의 금지' 등 10개 항목으로 이루어져 있는데, 이에 준거하여 소프트웨어 라이선스에는 'OS인정 마크'가 부여된다.

- 오픈소스 소프트웨어 라이센스 가이드
  http://wiki.kldp.org/wiki.php/OpenSourceLicenseGuide#s-1

② 감정시 오픈 소스의 제거

오픈 소스는 공개적으로 사용할 수 있는 프로그램이므로 유사도 감정시 이 부분은 반드시 제외하고 유사도를 산정하여야 한다.

③ 오픈 소스 검색 기관

- 한국저작권위원회에서 국가 및 공공분야에서 개발한 SW에 대한 무료 라이선스 검사서비스 '코드아이(CodeEye) 서비스'를 제공한다. 코드아이 서비스는 사용자가 저작한 소스코드를 위원회가 보유하고 있는 오픈소스 SW 데이터베이스(DB)와 비교 분석해 오픈소스 SW 사용에 따른 명확한 라이선스 정보 및 가이드를 제공하는 시스템이다.
- 영리 목적으로 운영되는 블랙덕(주)에서 일정 비용을 받고 오픈 소스를 검색해 준다.
  http://www. blackducksoftware.co.kr/

IV. 유사도 감정의 비교(Comparison)

1) DoubleKiller

DoubleKiller는 고소인측과 피고소인측 파일에서 동일한 데이터를 갖는 중복 파일 검색 삭제 프로그램이다. DoubleKiller는 옵션(날짜, 파일명, 내용 등)에 따라 서브폴더내 중복된 파일을 검색하고 삭제해주는 기능을 하므로 내용이 동일한 사운드 파일을 효과적으로 검색할 수 있다.

2) SIM(Software Similarity Tester)

SIM은 암스테르담의 Vrije 대학에서 개발되었으며 C, Java, Pascal, Modula-2, Miranda 및 자연어로 작성된 대규모 파일들의 유사도를 측정할 수 있다.

SIM 옵션은 다음과 같다.

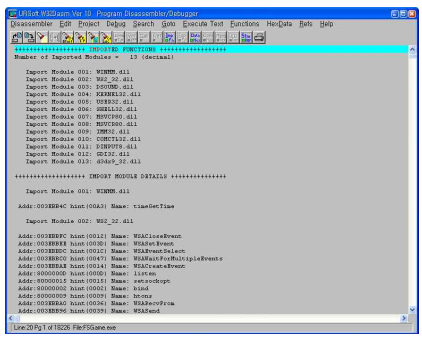
-s : 하나의 파일내 토큰끼리는 유사성 비교안함
-r 20 : 유사·동일하다고 인정하기 위해 연속적으로 일치해야 하는 최소 토큰 수
-p : 유사율을 %로 출력
-o file : 비교결과를 파일로 저장

3) exEyes

exEyes는 소스코드를 비교하는 주요기능 외에 함수호출관계도 보기, 클래스 다이어그램 보기, 변수 리스트 보기와 같은 다양한 기능을 가지고 있어 프로그램간의 실질적 유사성을 판단하는데 유용한 도구이다. exEyes는 파일을 비교하기 위하여 우선 비교대상 파일에서 의미 없는 공백을 제거하고, 소스코드를 라인단위로 병합 또는 분리한 후 토큰으로 파싱한다. 또한 모든 소스코드에 대한 비교쌍을 생성하여 비교하므로 파일명이 다르거나 일부 소스코드가 다른 파일이 존재하는 경우에도 검출이 가능하다.

4) W32Dasm

W32Dasm은 역어셈블러로 PE 헤더 및 역어셈블 코드를 보여주며 어셈블리어 수준의 디버깅 기능을 제공한다. 본 감정에서는 W32Dasm을 이용하여 PE 헤더의 Import/Export 참조 함수명을 추출한 후 상호 비교한다.



[그림 3.6] PE 헤더의 Import 함수 정보

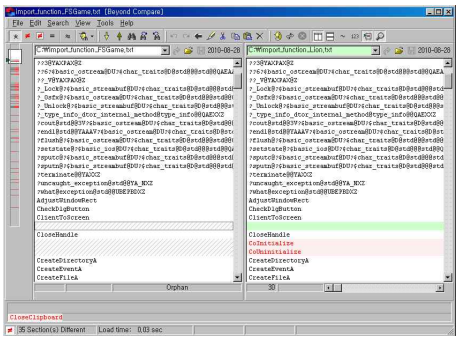
[그림 3.6]는 포더사커 실행파일 FSGame.exe의 Import 함수 정보를 보여주고 있으며 Export 함수는 없었다.

5) IDA pro

IDA pro의 Functions windows 보기 기능을 사용하면, DLL의 이진코드에서 역어셈블된 함수명을 추출할 수 있다. IDA pro의 Strings windows 보기 기능을 사용하면, 실행파일에 포함된, 텍스트 형태의 데이터를 추출할 수 있다. 이진코드에 들어 있는 텍스트 형태의 데이터로는 에러 메시지, 정보제공 메시지, 컴파일시 포함된 일부 함수명, 헤더에 포함되어 있던 미리 정의된 상수명(pre-defined constant name) 등이 해당된다. 이러한 스트링 정보는 사용자와의 인터페이스 측면에 많이 관여하기 때문에, 두 실행파일의 유사도를 비교하는데 직접적인 측정 기준으로 사용될 수 있다.

6) Beyond Compare

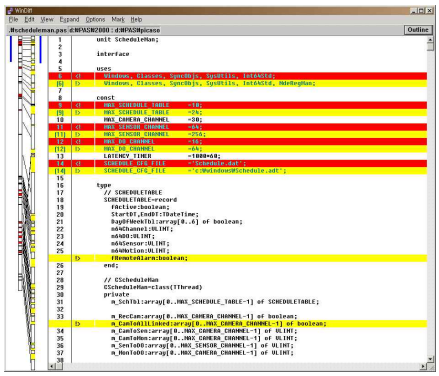
Beyond Compare를 이용하여 원본파일(FSGame.exe)을 기준으로 비교하였다. 비교결과 <그림 3.7>와 같이 원본파일의 전체 Import 함수명 개수는 294(=293+1)개 이고 그중에서 동일한 함수명 개수는 293개로 유사도는 99.66% 이다.



[그림 3.7] Import 함수명 비교

7) Windiff

WinDiff는 에서는 두 파일간의 비교 결과를 표시하는 기능이 있는데 이를 활용하였다.



[그림 3.8] WinDiff 프로그램의 파일 비교 결과 예

[그림 3.8] A와 B에 모두 있는 파일의 비교 결과 화면이다. 그림에서 흰 배경 부분은 동일한 부분, 빨간 배경 또는 검은 배경은 A의 파일에만 있는 부분, 노란 배경 또는 회색 배경은 B의 파일에만 있는 부분을 표시한다. 화면 왼쪽은 전체적인 동일성과 차이성의 양상을 나타내며, 왼쪽과 오른쪽의 수직 방향 바에서 서로 연결된 부분은 같은 부분이라는 것을 표시한다.

V. 결 론

본 논문에서는 유사도 감정에서 추상화란 소스 프로그램을 어떻게 분석할 것인가에 대한 구체적인 분석 방법 및 기준 제시와 감정시 여과하여야할 항목 및 내용에 대하여 설명하였다.

향후 연구 방향은 각 항목별 연구를 전문화하여 효율적인 감정 지침서를 만드는데 기여를 하고자한다.

참 고 문 헌

[1] 한국저작권위원회 홈 페이지, 정보자료 http://www.copyright.or.kr/info/law/precedent_list.do
[2] 방효근외 2명, “소프트웨어 복제 감정을 위한 유사도 및 복제도 평가방법에 관한 연구”, 한국소프트웨어감정평가학회 제2회 추계 학술발표대회 논문집, pp.9-15, 2003. 11.
[3] 신동명외 2명, “S/W하자 및 완성도 감정을 위한 평가방법 및 표준화모델에 관한 연구”, 한국소프트웨어감정평가학회 춘계 학술발표대회 논문집, pp.91-97, 2004. 5.
[4] 이성훈, 조광문, "소프트웨어 개발에서의 분쟁조정을 위한 표준화 방안" 한국콘텐츠학회 춘계 종합학술대회 논문집 , pp.406-409 , 2005
[5] Conformance Test Suite Software, http://www.itl.nist.gov/div897/ctg/software.htm
[6] Conformance Testing Procedure, http://www.controlnet.org/05_conf/05_ct_faq.htm

저 자 소 개

전 병 태

2001년 2월 : 고려대학교 박사
1989 ~ 1896년 : 한국과학기술 연구원 (KIST) 연구원
2004년 2월 ~ 현재 : 국립 한경대학교 웹정보공학과 교수
1992년 5월 : IR52 장영실 상 수상(과기부 장관상)
2007년 5월 ~ 2010년 5월 : 한국전자통신연구원 초빙연구원
2003년 8월 ~ 현재 : 한국저작권위원 SW감정전문위원
2009년 2월 ~ 현재 : 한국지식정보학회 이사
2004년 2월 ~ 현재 : 한국소프트웨어감정평가학회 부회장

<관심분야: 영상처리, 멀티미디어 영상처리>