

전자책 장애인 접근성 자동 검증도구용 라이브러리 연구

김현영[†], 임순범^{**}

Accessibility Automatic Inspector Library for EPUB and its Components

Hyun-Young Kim[†], Soon-Bum Lim^{**}

ABSTRACT

We have heard about several keywords of the fourth industrial revolution such as automation, artificial intelligence and connectivity. And electronic publishing systems and ebooks will enable us to experience a new industrial revolution without discrimination. On that level, it is necessary to study a method of automatic accessibility tool for electronic publishing. There are some studies on accessibility standard specifications, accessibility guidelines, and accessibility certificate criteria. but there is little research about automatic inspection tools in electronic publishing area. This paper proposed a library for accessibility inspector of EPUB and its components based on IDPF EPUB accessibility guidelines and TTA accessibility certificate criteria for reading disabled people. After implementing the library that can be connected to SIGIL, which is the most widely used in Korea, we verified its functionality and its coverage with Korean commercial EPUB files.

Key words: ebook Accessibility, Automatic Checker, Accessibility Evaluation Items, Verification

1. 서 론

4차 산업혁명을 맞이하여 정보접근성이 그 어느 때보다 중요해졌다는 주장이 제기되고 있다[1]. 지난해 말에 개최된 ‘2016 정보접근성 기술 컨퍼런스’에서 대한민국 ICT 기술 발전 속도가 세계 1위지만 이는 외형적 성장에 그치며 전체 인구 5%에 해당하는 250만명 장애인이나 고령층 등 취약계층이 정보를 고루 누리는 환경을 준비해야 한다고 했다.

널리 보급된 웹이나 모바일 접근성과 달리 전자책과 다른 신기술 영역은 접근성 확보한 정보 제작 및

유통 비율이 다른 영역에 비해 낮음으로 취약계층 정보 이용이 더욱 어려운 실정이다. 정보화 시대에서 정보 격차는 교육·소득 등 다른 격차를 더욱 벌어지게 하였으며 제4차 산업혁명 시대가 도래하면 이런 격차는 더 커질 것이라고 경고했다[2].

이에 따라 스마트 디바이스 상에서 정보 전달을 위하여 웹 및 모바일 앱 접근성 연구가 많이 진행되어 왔다. 웹 접근성 1.0 및 2.0 표준 발표에 이어 2.1 개정 작업이 예정 되어 있고, 모바일 접근성 1.0 및 2.0 표준 제정도 이루어 졌다. 이들을 기반으로 하여 웹 접근성을 준수하는 콘텐츠 제작을 위해 저작도구

※ Corresponding Author: Soon-Bum Lim, Address: (04310) 100, Chungpa-47gil, Yongsan-gu, Seoul, Korea, TEL: +82-10-5329-9919, FAX: +82-2-718-2337, E-mail: sblim@sm.ac.kr

Receipt date: Jan. 4, 2017, Approval date: Jan. 26, 2017
[†] Dept. of IT Engineering, SookMyung Women's University (E-mail: corolla.kim@gmail.com)

^{**} Dept. of IT Engineering, SookMyung Women's University

※ This research was partly supported by Basic Science Research Program through the National Research Foundation of Korea (NRF) funded by the Ministry of Education, Science and Technology (NFR-2015R1D1A1A09060170).

와 검증도구, 인증도구 등등의 웹 접근성 생태계를 위한 표준과 도구들이 존재한다. 또한 책이라는 메타포 형태로 정보를 전달하려는 전자책에 대한 접근성 연구도 국내외적으로 다양하게 진행되고 있다.[3]

전자책 접근성 부분을 좀 더 자세히 고찰하면, 국내외 전자책의 접근성 표준 규격들이 있으나 전자책 접근성 생태계를 위해 필수적인 표준 규격 준수 검증도구들은 아직 상용화된 것이 없다[4]. 현재 전자책 접근성의 검증 방법은 전자책 구조를 분해하여 코드 단위로 수동 점검하는 것뿐임으로 전자책 접근성 표준 규격에 기반 한 자동 도구의 개발이 필요한 시점이다. 또한 접근성 검증도구는 시장에서 사용하고 있는 전자책 저작도구와 연계하여 동작하여야만 전자책 제작의 편리성이 확보될 수 있고 실제 제작시간 단축, 제작비용 단축 등의 효과도 얻을 수 있음으로 여러 저작도구와 연동할 수 있는 라이브러리 형태로 개발하는 것이 바람직하다.

본 논문에서는 국제 전자책 표준 포맷인 EPUB3의 접근성 표준 규격 준수 여부를 검증하는 자동 검증 라이브러리를 제안하며, 자동 검증의 대상 항목은 국내 표준 규격에 준하여 48개 항목으로 진행하였다. 또한 한 라이브러리 안에서 EPUB3 구조체와 EPUB3 구성요소 각각에 대한 접근성 규격을 검증하며, 검증 오류 영역에 대한 명시와 검증오류의 원인 제시 등으로 실 제작시점에 유용하게 활용할 수 있는 기능을 구현하였다.

제안한 라이브러리의 검증을 위하여, 국내시장에서 가장 많이 사용하고 있는 저작도구와 연결하여 자동검증 항목 전체 48개의 오류를 가진 EPUB3 샘플파일을 사용하여 실제 48개 오류 항목과 오류 위치가 모두 검출되는 결과를 확인하였으며, 또한 국내 대표 출판사 EPUB3 파일 30개를 사용하여 일반적인 상용 전자책에 대한 접근성 검증능력을 점검하였다.

2. 접근성 관련 연구

2.1 접근성 표준 규격

국외 접근성 표준은 W3C(World Wide Web Consortium) 및 IDPF(International Digital Publication Forum), DAISY Consortium, DIAGRAM Center 등에서 작업한 것들이 있으며, 특히 전자책 관련 표준 규격을 정의하는 IDPF의 규격문서는 “EPUB Acces-

sibility Guidelines[5]” 및 이를 기반한 표준화 작업 [6]이 있으며, 이들은 W3C “Web Content Accessibility Guidelines (WCAG) 2.0 [7]” 및 “Accessible Rich Internet Applications (WAI-ARIA) 1.0 [8]”에서 정의한 웹 접근성 규격 중 HTML 및 CSS, JavaScript 웹 문서 규격을 수용하여 정리되었다.

또한 W3C 규격들과 IDPF 규격들 간의 차이점은 전자책 구조화 및 네비게이션(Navigation), 텍스트 읽어주기(Text-to-Speech), 논리적 읽기 순서 지정 등 전자책에 특화된 부분은 별도 정의되어져 있다는 것이다.

이외 표준 규격은 접근성 음성도서인 DAISY Consortium의 Digital Talking Book에서 정의한 접근성 요소들[9]이 있다. 해당 표준규격에서는 콘텐츠 접근성 지침, 저작도구 접근성 지침, 사용자 에이전트(사용자용 어플리케이션) 접근성 지침으로 나누어져 있으며 콘텐츠 접근성 지침 항목들은 IDPF 규격과 거의 유사하다.

국내 접근성 표준 규격은 한국정보통신기술협회(TTA)에서 정의한 “TTAK.KO-10.0905-Part1”[10] 및 “TTAK.KO-10.0905-Part2”[11] 등이 존재하며 154개 항목의 접근성 저작지침들과 139개의 인증기준들이 있다. 특히 TTAK.KO-10.0905-Part2에서 정의한 139개 인증기준들은 59개의 자동진단항목과 80개의 수동진단항목을 구분하여 기준을 정의하였으므로 이번 연구에서도 이를 기반으로 전자책 접근성 자동검증 라이브러리를 연구하는 것이 타당하다.

2.2 EPUB 자동검증 라이브러리

전자책 접근성 관련된 라이브러리는 시장에 없으나 EPUB 규격준수 검증에 대한 라이브러리는 IDPF에서 EpubCheck[12]으로 발표하였다. 이 라이브러리는 EPUB 2.0 및 3.0 규격 모두를 지원하며, 각 버전별 구조적 검증 및 EPUB 규격 요소인 OCF(Open Container Format) 컨테이너 구조, OPF(Open Packaging Format) 및 OPS(Open Publication Structure) EPUB publications 3.0 마크업, 내부참조모델의 일관성들을 자동 검증하는 역할을 수행한다.

본 논문에서 진행하는 라이브러리 역시 EPUB 2.0 및 3.0 규격에 적용할 수 있어야 하며, 콘텐츠 정의영역 뿐만 아니라 EPUB 구조에 대한 접근성 검증이

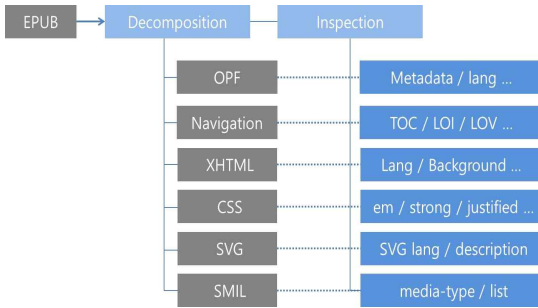


Fig. 1. Process for accessibility inspection algorithm.

이루어져야 함으로 EPUB 자동검증 라이브러리의 구조분해 및 내부참조 알고리즘을 활용하여 진행하여야 한다.

3. 접근성 검증 라이브러리 구현

3.1 제안한 알고리즘의 개요

본 논문에서는 EPUB과 그의 구성물에 대한 접근성 자동 검증을 진행함으로 1단계에서는 EPUB 구조를 분해하고 각 분해한 요소별, 상호참조 관계별 접근성을 검증한다. 2단계에서는 각 구성요소 자체적인 접근성을 검증한다. Fig. 1은 자동 검증 알고리즘의 전체적 흐름을 보여준다.

자동검증 라이브러리를 구현 위해서는 일단 EPUB 구조를 분해해야 함으로 EPUB ZIP 구조 처리하는 라이브러리가 필요하며, 또한 XML/XHTML, CSS,

SMIL, SVG 등의 구문을 이해하고 이의 접근성 요소를 검증하여야 한다. 따라서 본 논문에서는 이 들을 위한 기본적인 작업을 공개 라이브러리를 활용하여 구현하였다. 한편 자동 검증 라이브러리가 실제 전자책 저작도구나 전자책 관련 도구들과 연동하기 위하여 검증된 결과물을 연계할 필요가 발생한다. 이에 검증된 결과물, 오류 발생한 구문과 위치, 오류 수정용 설명문구 등에 대한 전달이 별도의 처리 없이 쉽게 진행하기 위하여 JSON 형식으로 처리하였다.

EPUB 구조 이해는 EpubCheck 라이브러리에서 구현한 방법을 참고하여 OPF 구조정의 파일에 대한 파싱을 통해 진행하였다. 또한 이를 통해 구조정의 파일이 포함하고 있는 각 구성요소의 파일 위치, 구성요소간의 포함관계를 파악하게 실질적인 EPUB 구성요소들을 분해하였다.

3.2 자동검증 항목의 추출 및 해당 샘플 생성

구조 분해 후 실질적인 자동 검증 분석 모듈은 TTA 접근성 인증의 자동진단 항목들에 기초하여 구현하였다. 앞서 확인한 TTA 접근성 자동진단 항목은 모두 59개이다. 이 중 Fig. 2에서 알 수 있듯이 뷰어의 접근성 기능 영역, EPUB 시멘틱 용어 규정, ARIA role 속성, EPUB 의미구조 정의 등 콘텐츠의 접근성과 무관한 항목을 제외하였다. 따라서 전자책 콘텐츠에 대한 자동검증 라이브러리에서 구현 가능한 48개 항목에 대해 진행하였다.

SVG 이미지가 포함된 텍스트 콘텐츠의 언어는 항상 보조 기술(AT)이 제공된 내용을 포함하고 생성할 수 있도록 설정하여야 한다. svg 요소에 포함된 xml:lang 속성은 이미지 안에 있는 텍스트 콘텐츠의 기본 언어를 식별한다.

또한 xml:lang 속성을 각 구성요소에 사용하여 svg 요소의 기본(default) 언어가 아닌 다른 언어의 텍스트가 포함되어 있다는 것을 나타낸다.

제목과 설명(Title and Description): SVG 이미지는 반드시 제목과 설명이 있어야 한다.

출판물에서 중요한 의미를 가진 SVG 이미지는 보조 기술(AT)을 사용하는 독자들을 위해 alt(alt)과 desc(desc)를 포함하여야 한다. desc에는 설명은 이미지에 포함된 정보에 상응하여야 한다. 간단한 이미지나 조금 더 복잡한 이미지를 위해 alt와 desc 설명을 사용한다.

오른 SVG 이미지가 장식용이라면 제목과 설명을 포함시킬 필요는 없다. 장식용 목적으로 하는 SVG 요소는 ARIA role 속성을 통해 식별하여야 한다. 그러나 SVG 이미지 내의 모든 중요한 그래픽 요소들은 이미지의 전반적인 접근성을 향상시키기 위해서 제목과 설명을 포함하여야 한다.

구성요소(Component): SVG 이미지는 재사용 가능하도록 구성하여야 한다.

SVG 이미지의 이해를 돕기 위해 그래픽 구성요소는 재사용한다. SVG는 defs 요소를 포함하여 공통 구성요소를 정의한다. 새로운 구성요소를 도입하기 위해서 필요에 따라

하는 대신 문장을 참조하기 위해 ref 요소를 사용하여야 한다.

스타일링(Styling): 그래픽 구성요소에 직접적으로 스타일 속성을 추가해서는 안된다.

가능하면, 별도의 스타일시트에서 그래픽 구성요소를 위한 스타일 속성(style property)을 정의한다. 스타일을 분리하면 기본 표현방식에 문제가 생길 때 사용자 지정 스타일을 적용할 수 있도록 해야 한다.

그래픽 구성요소에 직접적으로 style 속성을 추가하지 않아야 한다. 형식을 변경하고자 하는 것과의 기능을 변경할 수 있기 때문이다. 텍스트를 삽입할 때마다, 글꼴 래핑 이미지 제거 세로 흔들리기 알고 구분자는 영도 확인한다.

관련된 추가사항은 EPUB 스타일 시트(EPUB Style Sheets) 보기를 참조한다.

자항 7, 상호작용 (인접혁신)	
항목지항 28, 미디어오버레이(Media Overlay): 출판물 내부 관련 콘텐츠 문서는 스스로의 언어와 오버레이를 가져야 한다.	

원문항목	
자가진단	
7.6	package 문서의 manifest 상에서 media-type="application/smil+xml"을 갖는 항목의 ID가 다른 항목의 media-overlay 값으로 참조되어야 한다
7.7	package 문서에서 미디어 오버레이와 연결된 content 문서는 manifest로 기술된 media-overlay 값이 있어야 하고, 그 같은 media-

Fig. 2. Automatic inspection items for accessibility certificate criteria.

완성되어진 접근성 검증 라이브러리의 기능 완성도 및 충실도를 확인할 수 있도록 라이브러리에서 검증한 48개의 모든 오류를 포함한 샘플 EPUB을 제작하였다. 샘플 EPUB의 제작은 IDPF EPUB accessibility guideline에서 각각의 해당 오류 설명 구문으로 제시되어진 바르게 사용된 예제와 잘못 사용된 예제를 모두 포함하도록 구성하였다. 라이브러리에서 샘플 EPUB을 검증하면 바르게 사용된 부분은 추출되지 않고, 잘못 사용된 부분만 추출되는 것을 확인하면 기능 완성도와 충실도를 점검할 수 있게 된다.

4. 실험 결과 및 고찰

4.1 저작도구와의 연결

국내에서 EPUB 저작도구로 가장 많이 활용되고 있는 SIGIL[13]과 연계하여 라이브러리 기능을 점검하였다. SIGIL 윈도우 버전에 본 논문의 접근성 자동 검증 라이브러리를 연결하여 48개 항목 오류를 모두 포함하고 있는 샘플 EPUB에서 48개 오류 검증을 모두 진행하는 것을 확인함으로써 기능의 완성도를 점검하였다. Fig. 3에서 SIGIL과 연동한 라이브러리 화면을 확인할 수 있다.

4.2 상용 전자책에서의 검증

본 연구에서 가공한 샘플 EPUB 뿐만 아니라 국내 대표 출판사에서 제작한 전자책을 활용하여 자동 검증 라이브러리의 실무 적용성을 점검하였다. 무료 전자책 서비스 사이트를 통해 체험용으로 제공하는 문학동네, 황금가지, 위즈덤하우스, 민음사, 휴머니스트 등의 대표 출판사의 소설류 및 비소설류 EPUB 30개를 활용하여 점검하였다. SIGIL 윈도우 버전에 본 논문의 접근성 자동 검증 라이브러리를 연결하여 실제 출판사에서 제작한 상용 전자책에서 오류 검증 기능을 확인하였다.

국내 전자책 시장이 아직 EPUB3 멀티미디어, 표, SVG, TTS 등 다양한 항목으로 구성되지 않았기에 대부분의 오류는 텍스트 및 스타일 처리 영역에서 발생할 것이라 예상하였고, 실제 검증에서도 한영혼용 문구에서의 언어정의, 그림에 대한 대체텍스트 정의, 장이나 절 제목에 대한 h1/h2 등의 제목 태그 지정, 양끝정렬 오류, 스타일 파일 분리 처리 등에서 발생하였고 자동 검증 기능이 정상적으로 동작하는 것을 확인할 수 있었다. Table 1은 30개 상용샘플에서 접근성 자동 검증을 통해 발견한 오류별 발견 개수 및 이에 대한 전체 개수 대비 비율이다.

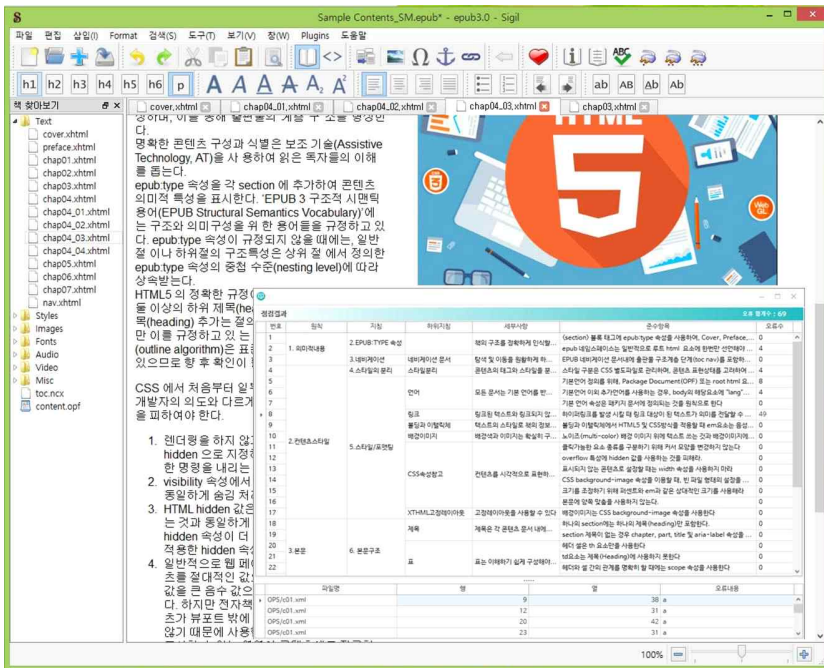


Fig. 3. Sigil window connected to accessibility automatic inspection library.

Table 1. Accessibility errors from 30 commercial EPUB files

Detected Items		# of errors	Ratio
1	The fig caption element must be the very first or last in the figure	4	0.08
2	Each section should have a numbered heading. Each heading element must represent a single heading.	350	7.37
3	Use relative sizes such as percentages and ems to facilitate scaling	18	0.38
4	Avoid justifying text, as the uneven spacing that occurs between words can reduce the readability for some people.	25	0.53
5	The CSS properties for bolding and italics should be used whenever the use of bolding and italics is presentational	215	4.53
6	Separating style from markup is about keeping CSS in a separate file from your markup	327	6.88
7	Hyperlinked text should be discoverable by more than just its color. Hyperlinked text may have an alternate text like a tag with title attribute.	152	3.20
8	Images that are central to the understanding of a publication must always include a text alternative in their alt attribute.	325	6.84
9	When using the epub:type attribute in a content document, the epub namespace must be declared on the element containing the attribute, or on one of its ancestors.	591	12.44
10	The language attributes should identify the language of the text content of that element to identify a change in language.	915	19.26
11	To define the default language for an XHTML document, the lang and xml:lang language attributes need to be attached to the root html element. The language code defined in these attributes should match one of the language codes specified in a dc:language element in the package document's metadata section.	1828	38.48
Total		4750	100.00

5. 결 론

본 논문에서는 전자책 접근성 표준안에 기반하여 자동 검증 라이브러리를 제안하였다. 이 라이브러리는 EPUB 구조적 분석과 더불어 구성요소에 대한 48개 접근성 항목 기반의 자동 검증을 진행한다. 또한 접근성 오류를 검출하는 것과 더불어, 오류 영역에 대한 정확한 명시(코드 레벨에서의 행수, 열수), 오류에 대한 수정 가이드(수정/대체할 명령어 예시) 등을 제공하여 자동 검증된 오류를 바로 수정할 수 있도록 안내해 준다.

종이책 시장과 함께 성장하는 전자책에 접근성 점검을 도와줌으로써 접근 가능한 전자책 제작에 소요되는 시간과 비용을 절감하는데 도움이 될 수 있을 것으로 예상하며, 실제 국내에서 여러 가지 다양한 저작도구들과 연동하여 활용할 수 있도록 라이브러

리 형태로 구현하였다. 기능의 검증은 가장 점유율이 높은 SIGIL에서 진행하여 오류 검증 능력을 확인하였다.

TTA 전자책 접근성 인증기준에서 언급되었듯이 전자책에 대한 독서 장애인들의 이해를 높이려면 현재 상용 전자책보다 다양하고 많은 멀티미디어의 멀티 모달 방법적 지원과 활용이 필요한 상태이다. 그러나 한편으로 멀티미디어 및 멀티 모달 비율이 높아 질수록 지금보다 더 복잡한 접근성 기능이 구현되어야 함으로 전자책 접근성 검증도구의 지속적인 업그레이드가 필요할 것이며, 또한 이들에 대한 검증을 수동으로 진행하여야 하는 문제도 증가할 것임으로 자동 검증도구의 자동화 능력을 확장하기 위해 자연어 처리와 유사한 EPUB 구문분석, 의미분석 기반의 기계학습적 기법을 적용한 연구로 확장할 필요가 있다.

REFERENCE

[1] C.Y. Son, "Direction of Information Accessibility Enhancement Policy," *Proceeding of 2016 Information Accessibility Technology Conference*, pp. 1-12, 2016.

[2] S.S. Noh, "Trends and Countermeasures of Web Accessibility International Standard Revision," *Proceeding of Information Accessibility Technology Conference*, pp. 13-35, 2016.

[3] K.H. Lee, T.E. Kim, J.W. Lee, S.B. Lim, "A Design of Mobile e-Book Viewer interface for the Reading Disabled People," *Journal of Korea Multimedia Society, Vol. 16, No. 1*, pp. 100-107, 2013.

[4] H.Y. Kim, "eBook Accessibility Standard," *Korea Telecommunication Technology Association TTA Journal*, Vol. 166, pp. 76-81, 2016.

[5] IDPF "EPUB 3 Accessibility Guidelines." <https://idpf.github.io/a11y-guidelines/> (accessed Dec. 30 2016).

[6] IDPF, EPUB Accessibility 1.0., <http://www.idpf.org/epub/a11y/accessibility.html> (accessed Dec. 30, 2016).

[7] World Wide Web Consortium, "Web Content Accessibility Guidelines (WCAG) 2.0, <https://www.w3.org/TR/WCAG20/> (accessed Dec. 23, 2016).

[8] World Wide Web Consortium, Accessible Rich Internet Applications (WAI-ARIA) 1.0, <https://www.w3.org/TR/wai-aria/> (accessed Dec. 23, 2016).

[9] ANSI/NISO Z39.86-2005 (R2012) Revision of ANSI/NISO Z39.86-2002, Specifications for the Digital Talking Book, <http://www.daisy.org/z3986/2005/Z3986-2005.html?q=z3986/2005/z3986-2005.html> (accessed Jan. 02, 2017).

[10] Telecommunication Technology Association ePublishing PG, *Accessibility Guidelines for Reading Disabled People-Part 1: Authoring Rules*, 2016.

[11] Telecommunication Technology Association ePublishing PG, *Accessibility Guidelines for Reading Disabled People - Part 2: Certificate Criteria*, 2016.

[12] IDPF, EpubCheck, <https://github.com/IDPF/epubcheck>, (accessed Jan. 02, 2017).

[13] SIGIL, Sigil-Ebook, Sigil-0.9.7, <https://github.com/Sigil-Ebook/Sigil/releases> (accessed Nov. 01, 2016).



김 현 영

1988년 서울대학교 수학교육학과 (학사)
 1990년 한국과학기술원 수학과 (석사)
 2000년~2013년 (주)다우인큐브 디지털미디어 사업본부 (이사)

2014년~현재 숙명여자대학교 IT공학과 박사과정
 관심분야: 전자출판, 이러닝, 기계학습, 감성공학



임 순 범

1982년 서울대학교 계산통계학과 (학사)
 1983년 한국과학기술원 전산학과 (석사)
 1992년 한국과학기술원 전산학과 (박사)

1989~1992년 (주)휴먼컴퓨터 창업 (연구소장)
 1992~1997년 (주)삼보컴퓨터 프린터개발부 부장
 1997~2001년 건국대학교 컴퓨터학과 교수
 2001년~현재 숙명여자대학교 IT공학과 교수
 관심분야: 컴퓨터 그래픽스, 웹/모바일 멀티미디어 응용, 디지털 방송, 전자출판(폰트, 전자책, XML 문서), User Interface