읽기 매체의 다양성과 흥미도를 고려한 가독성 측정

김아영[O], 박성배, 이상조
경북대학교, 컴퓨터학부
{aykim, sbpark}@sejong.knu.ac.kr sjlee@knu.ac.kr

Revisiting Readability: An aspect of the reading environments and perspectives

A-Yeong Kim[O], Seong-Bae Park, Sang-Jo Lee
Kyungpook National University, Department of Computer Science and Engineering

요 약

가독성이란 글을 읽음에 있어 얼마나 쉽게 쓰여졌는가, 얼마나 흥미로운가의 정도를 나타낸다. 기존 연구에서는 가독성을 측정함에 있어 인쇄물에 한정하고 글의 구성 관점에서만을 반영하였다. 하지만, 글은 인쇄물뿐만 아니라 모바일, 태블릿 등 다양한 읽기 매체를 통해 읽어질 수 있으며 글의 흥미에 따라 가독성이 달라지므로 기존의 연구 방법으로는 적절히 반영할 수 없다. 본 논문에서는 읽기 매체에 따라 가독성이 어떻게 달라지는지, 가독성을 측정함에 있어 관점들이 어떻게 반영되는지를 살펴보고자 한다. 서로 다른 읽기 매체들을 사용하여 동일한 뉴스 기사에 대해 가독성을 측정하였으며, 측정된 결과를 사용하여 각 읽기 매체가 가독성를 측정하는 요소들 중 어떤 요소에 큰 영향을 받는지를 살펴본다. 또한, 가독성을 측정함에 있어 글의 구성뿐만 아니라 흥미 관점을 추가하였으며, 이로부터 가독성의 요소들이 글의 구성 관점과 흥미 관점에서 어떠한 차이점을 보이는지 살펴본다.

주제어: 가독성, 읽기 매체, 가독성 관점

1. 서론

가독성(Readability)이란 글을 읽음에 있어 얼마나 쉽게 쓰여 있는가(well-written)[1], 얼마나 흥미로운가(interesting)를 의미한다[2]. 가독성은 글을 읽음에 있어 난이도를 도입하여 학생들의 수준에 맞는 글을 가르치고자 측정하게 되었다[3]. 일반적으로 가독성은 가독성을 결정하는 각 요소(factor)들에 의해 측정된다[4,5,6,7]. 자연언어처리 관점에서 가독성을 결정하는 요소로는 글의 길이, 명사구 또는 동사구의 비율, 대명사의 비율 등이 존재한다[4]. 최근 들어 교육적인 목적뿐만 아니라 다양한 매체에서 가독성을 구하거나 가독성에 영향을 주는 새로운 요소를 제안하는 연구들이 활발히 진행되고 있다[4,5,8].

기존의 자연언어처리 측면에서의 가독성 측정 연구들은 주로 인쇄물에서 집중되어 왔다[4,5,6,7]. 하지만, 최근 들어 읽기 매체가 다양해짐에 따라 사용자는 인쇄물뿐만 아니라 스마트폰, 태블릿과 같은 휴대용 기기를 사용하여 글을 읽을 수 있게 되었다. 각 읽기 매체들은 매체들 마다 특성이 존재하므로 동일한 글이라도 가독성이 달라질 수 있다. 또한 기존의 방법들은 글의 구성(well-written) 관점만을 고려하여 가독성을 측정하였다[2,4,5]. 하지만 가독성은 글의 구성뿐만 아니라 흥미도에도 영향을 받을 수 있기 때문에 가독성을 측정함에 있어 각 관점을 고려할 필요가 있다.

본 논문에서는 읽기 매체의 다양성 및 가독성의 다양한 관점을 고려하여 가독성을 측정하고 다음의 이슈들을 살펴보고자 한다. 1) 가독성을 결정하는 요소들은 인쇄물과 휴대용 기기에서 각각 다른 중요도로 적용 될 것인가? 2) 가독성을 측정하는 관점에 따라 가독성 측정에 영향을 미치는 요소들이 달라지는가?

본 논문에서는 뉴스 기사를 이용하여 인쇄물, 스마트폰, 태블릿에서 글의 구성, 글의 흥미를 반영하여 가독성을 측정한다. 같은 글에 대해 인쇄물과 휴대용 기기에서, 가독성을 결정하는 요소들과 가독성이 어떠한 상관관계를 가지는지 살펴본다. 또한 글의 구성을 나타내는 관점과 흥미도의 관점에서, 가독성의 요소와 가독성이 어떠한 상관관계를 가지는지를 관점마다 비교하며, 관점에 따라 서로 다른 특징을 살펴본다.

본 실험은 60개의 뉴스 기사에 대해 4명의 실험 대상자로부터 가독성을 측정하였다. 실험 결과, 첫 번째로 각 읽기 매체에 대해 상관관계가 유의미하게 보인 요소들이 서로 다름을 보이며, 이를 통해 읽는 매체마다 가독성의 차이가 존재함을 보였다. 다음으로 관점에 따라 영향을 주는 요소들의 차이가 존재하였다. 글의 구성 측면에서는 읽은 매체들이 서로 다른 요소들에 의해 영향을 받는 반면 글의 흥미 측면에서는 읽은 매체에 관계없이 요소들이 서로 비슷한 경향을 보였다. 이를 통해 글의 구성적인 측면에서는 매체마다 차이점을, 글의 흥미 측면에서는 매체가 무관한 경향을 보여주었다.

본 논문의 구성은 다음과 같다. 2장에서는 가독성 측정과 관련된 기존 연구들을 살펴본다. 3장에서는 가독성 측정 방법에 대하여 간략히 다루고, 본 논문에서 사용한 가독성을 결정하는 요소에 대해서 설명한다. 4장에서는 실험 및 결과를 분석하고, 마지막 5장에서는 결론을 다룬다.

2. 관련 연구

가독성을 측정하고자 하는 연구는 오래전부터 많이 진행되어왔다. 현재까지 가장 보편하게 널리 쓰이고 있는 측정 방법들은 주로 단어의 길이나 문장의 길이 또는 쉬운 단어의 빈도수를 바탕으로 측정하였다[6,7]. Flesch는 가독성을 "읽기의 쉬움" 이라 표현하고 말뭉치로부터 무작위로 100개의 단어를 선정, 단어의 평균 길이와 문장의 길이를 이용해 가독성 측정 공식을 도출하였다[6]. 이후 첫 번째 식을 보정하여 말뭉치의 특성을 더 잘 반영하도록 가독성 측정 공식을 확장하였다[7]. 이를 Flesch-Kincaid 식이라고 하며 현재까지도 교과서나 교육을 목적으로 한 책의 난이도를 측정하는데 쓰이고 있다. Dale은 3천개의 쉬운 단어 리스트를 미리 정의해두고, 단어가 포함된 문장의 평균 길이와 그 목록에 해당하는 단어의 백분율을 이용하였다[2]. 위 연구들은 단어의 길이나 문자의 길이 등 글의 표면적인 요소들만을 이용하여 가독성을 측정하였다. 하지만, 구문적인 요소를 반영하지 않기 때문에 정확한 가독성 측정을 할 수 없다는 한계를 보인다[5].

앞선 연구들의 단점을 보완하기 위하여 가독성을 측정하는 새로운 요소들을 추가하고 가독성의 측정 성능을 높이기 위한 요소들의 조합을 찾아내는 연구들이 진행되어 왔다[4,5]. Thomas et al.은 프랑스어를 외국어로 쓰는 사람들에게 맞는 새로운 가독성 측정 공식과 함께 기존 요소를 포함하여 46개의 요소들을 제안하였다[5]. 크게 어휘, 구문, 의미 그리고 프랑스어의 특성을 반영하는 요소와 같이 크게 4개로 분류하여 정의하였으며, 이를 바탕으로 가독성을 측정하였다. Pitler et al.은 글의 어휘, 구문, 담화를 나타내는 요소와 가독성의 관계를 살펴보았다[4]. 제안한 방법은 실험자가 글을 읽어 글에 대한 가독성 점수를 매긴 다음 그 점수와 가독성을 결정할 수 있는 요소간의 상관계수를 구했다. 상관계수를 이용하여 그 요소가 가독성을 판단하는데 얼마나 주요하게 영향을 미치는지 분석하였다. 그 결과, 문장 당 동사구의 평균 개수 및 글의 길이, 담화와 관련된 요소들이 가장 중요하게 영향을 끼쳤음을 보였다. 하지만 위의 방법들은 인쇄물 이외의 다른 읽기 매체들을 고려하지 않았으며, 글을 읽는 관점 또한 글의 짜임 및 구성(well-written)만을 고려하여 가독성을 측정하였다는 단점을 지닌다.

3. 가독성 측정

일반적으로 글 d가 주어졌을 때 가독성 R은 가독성을 결정할 수 있는 요소 x_i와 그 요소들의 중요도 α_i를 사용하여 계산될 수 있다[2,4,5,6,7]. 각 요소들이 가독성에 독립적으로 영향을 끼친다고 가정을 하면 가독성은 수식 (1)과 같이 선형 모델로 표현할 수 있다.

$$R(d) = \sum_{i}^{i} \alpha_i x_i \qquad (1)$$

각 요소의 가중치는 측정된 가독성 점수와의 선형 회귀(linear regression)를 통해 추정할 수 있다.

본 논문에서는 가독성을 결정하는 요소로 [4]에서 제안한 요소들 중 한국어에 적용할 수 있는 8가지를 선택하여 사용한다. 각 요소는 표면(Superficial), 어휘(Lexical), 구문(Syntactic), 문장의 연결성(Sentence Continuous)과 같이 4가지로 분류한다. 각 요소에 대한 설명은 다음과 같다.

표면적인 요소

표면적인 요소(Superficial factor)는 글의 구성적인 측면을 반영하고 있는 요소들이다[4,5]. 본 논문에서는 글의 길이(TL), 평균 문장 길이(SL), 문장 당 평균 어절 수(WS), 어절 당 평균 어절 길이(CW)를 사용하였다. 글의 길이는 글의 양을 측정하며 글이 지나치게 길게 되면 글의 흐름이 쉽게 끊길 수 있다. 글의 길이와 문장 당 평균 단어의 수는 수식어구의 영향을 받는다. 수식어구가 많으면 글의 길이와 문장 당 평균 단어 수도 같이 길어지거나 증가한다. 반대로 수식어구의 수가 적으면 글의 길이와 문장 당 평균 단어수가 같이 짧아지거나 줄어든다. 어절 당 평균 어절 길이는 전문용어와 같은 복합명사들을 반영할 수 있기 때문에 가독성에 영향을 주게 된다.

어휘적인 요소

어휘적인 요소(Lexical factor)는 글이 말뭉치로부터 자주 나오는 단어들로 구성되어있는지를 반영한다. 본 논문에서는 유니그램(unigram) 모델로부터 글에 쓰인 단어에 대한 확률을 구하였으며 식은 아래와 같다[4].

$$\prod_{w} P(w|M)^{C(w)}$$

여기서, $P(w|M)$은 말뭉치 M에 대한 단어 w가 나올 확률이고, $C(w)$는 글에서 단어 w가 나타나는 횟수이다. 위 수식에 대한 로그우도(log likelihood, LL)식은 다음과 같이 정의할 수 있다.

$$\sum_{w} C(w)\log(P(w|M))$$

글의 로그 우도의 값이 높으면 그 글은 주어진 말뭉치에서 주로 사용하는 어휘를 사용하고 있다는 것을 의미한다.

구문적인 요소

구문적인 요소(Syntactic factor)는 글의 구문적인 측면을 반영한다[4]. 문장에 구문적인 요소가 많을수록 문장을 복잡하게 만든다는 특성을 가지고 있다. 본 논문에서는 구문적인 요소로 문장 당 평균 명사구 수(NP), 문장 당 평균 동사구 수(VP)를 사용하였다. 명사구와 동사구는 많을수록 문장이 복잡해지지만 이해와 흥미를 높여주기 때문에 글의 흥미도를 반영할 수 있는 요소이다[9].

표 1 뉴스 기사에 대한 간단한 통계 자료

구분	최소값	최대값	평균값
글의 길이	107	546	326.23
문장 수	1	13	5.63
어절 수	3	47	17.48

문장의 연결성

문장의 연결성(Sentence Continuous)이란 주어진 문장들이 의미적으로 얼마나 잘 연결 되어있는지를 나타낸다. 본 논문에서는 문장의 연결성을 나타내는 요소로 문장 당 평균 대명사의 수(PRP)만을 사용하였다[4]. 대명사는 그 수가 많을수록 문장의 복잡도가 증가하지만 글의 흐름을 매끄럽게 하기 때문에 글의 구성적인 측면을 반영한다.

4. 실험

4.1 실험 데이터 및 가독성 채점 기준

실험을 위해 본 논문에서는 Naver에서 뉴스를 수집하였다. 6월 10일부터 6월 25일까지의 연예, 스포츠, 정치 3개의 카테고리에서 기사 60개를 수집하였다. 표 1은 실험에 사용된 기사 60개에 대한 통계적 정보를 보여준다. 가장 길이가 짧은 기사는 107자이며, 가장 긴 기사는 546자로 평균적으로 326.23자의 기사로 이루어져있다. 기사 한 개에 최소 1문장부터 최대 13문장으로 이루어져 있으며 평균적으로 5.63개의 문장으로 이루어져 있다. 한 문장은 최소 3개의 어절에서 최대 47개의 어절로 이루어져 있고 평균적으로 17.48개의 어절을 가지고 있다.

가독성 측정을 위해 학부생 2명에게 스마트폰(5인치), 태블릿(10.1인치), 인쇄물(A4용지) 총 세 가지 읽기 매체를 사용하여 60개의 기사를 읽게 하였다. 각 읽기 매체의 폰트 크기는 동일하게 하였다. 채점 기준은 두 가지로 다음과 같다.

▶ 주어진 기사는 구성이 좋다.

글이 얼마나 읽기 좋게 쓰여 있는지를 평가한다. 기존 연구에서 사용했던 주요 채점 방법이다. 본 논문에서는 주로 글의 짜임새나 흐름이 잘 흘러가는지를 중심으로 채점하였다.

표 2 채점 점수에 대한 통계 자료

구분	항목	최저 점수	최고 점수	평균 점수	분산
읽기 매체	스마트폰	2.5	5	3.767	0.432
	태블릿	2.5	5	3.867	0.295
	인쇄물	2	4.75	3.867	0.490
관점	글의구성	2	5	3.908	0.420
	글의흥미	1	5	3.550	0.735

▶ 주어진 기사는 흥미롭다.

주어진 기사의 흥미로움을 채점한다. 기존 연구에서는 가독성 측정에서의 글의 흥미도에 대한 언급은 하였지만 반영을 하지 않았다[4]. 하지만 가독성은 글의 구성뿐만 아니라 글의 흥미도와 함께 복합적으로 영향을 받기 때문에 본 논문에서는 글의 흥미도 또한 채점 기준으로 고려하였다.

점수는 5점(매우 좋다, 매우 흥미롭다)에서 1점(전혀 좋지 않다, 전혀 흥미롭지 못하다)까지 다섯 개의 등급으로 각 기준에 대하여 채점하였다. 최종적으로 각 기사마다 두 개의 기준에 대한 점수에 평균을 하여 가독성을 매겼다. 표 2는 채점 점수에 대한 통계 수치를 나타낸다. 글의 흥미를 제외한 나머지 항목은 최저 점수가 2점대였고, 인쇄물을 제외한 나머지 항목은 최고 점수가 5점 이었다. 평균 점수는 모든 항목들이 3점대 후반이었다. 카파 계수(Kappa value)는 0.626이다.

어휘적인 요소를 반영하는 로그우도를 계산하기 위해 말뭉치로 2013년도 1월 1일부터 2013년 9월 6일까지 Naver 기사를 수집하여 사용하였다. 수집된 전체 기사의 개수는 298,729개이다.

4.2 실험 결과

4.2.1 가독성 결정 요소와 읽기 매체와의 관계

표 3은 가독성 결정 요소와 가독성과의 상관관계를 읽기 매체별로 나타낸 것이다. 가독성 측정을 위해 사용한 8개의 요소에서 유의미한 값 중 상관관계가 높은 순서대로 나타내었다. 먼저 스마트폰의 경우 화면 크기의 제약 때문에 글의 길이(TL)가 크게 영향을 받음을 볼 수 있다. 이와 반대로 태블릿과 인쇄물은 화면의 제약이 없어

표 3 읽기 매체별 가독성 결정 요소와 가독성의 관계

스마트폰		태블릿		인쇄물	
요소	값	요소	값	요소	값
글의 길이(TL)	**-0.257**	**문장 당 평균 명사구 수(NP)**	**0.335**	**문장 당 평균 단어 수(WS)**	**0.315**
문장 당 평균 대명사 수(PRP)	**0.104**	**문장 당 평균 단어 수(WS)**	**0.276**	**문장 길이(SL)**	**0.294**
로그우도(LL)	-0.100	문장 길이(SL)	0.234	**문장 당 평균 명사구 수(NP)**	**0.264**
어절 당 어절 길이(CW)	-0.096	문장 당 평균 동사구 수(VP)	0.212	문장 당 평균 동사구 수(VP)	0.158

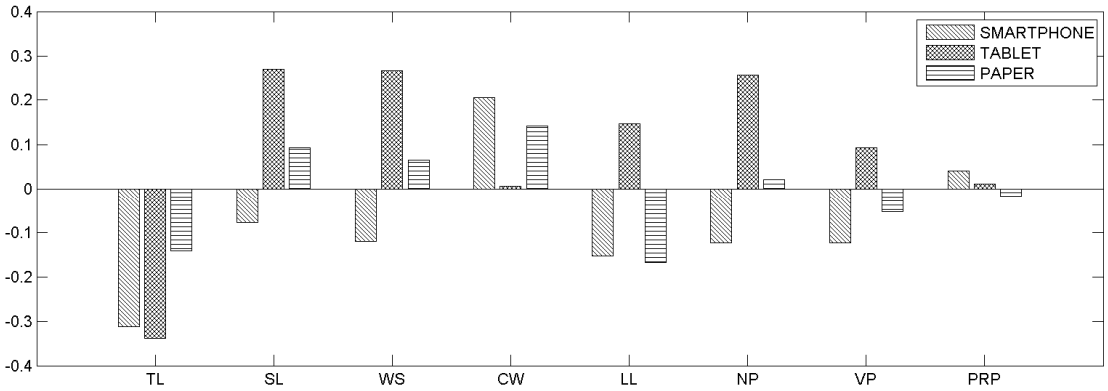

그림 1 글의 구성 관점에 대한 가독성과 요소간의 상관관계

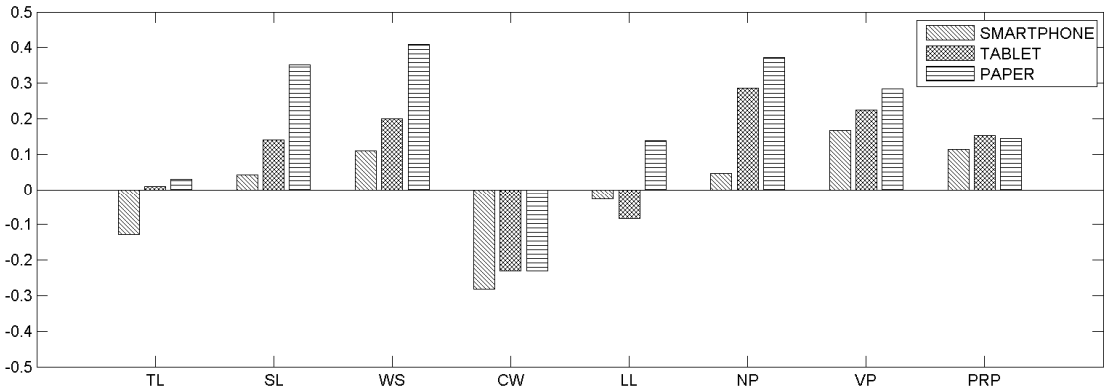

그림 2 글의 흥미 관점에 대한 가독성과 요소간의 상관관계

글의 내용과 관련된 요소들(NP, VP)에 영향을 받음을 볼 수 있다. 스마트폰에서의 대명사의 경우 수가 많을수록 문장의 복잡도가 높아지지만, 글의 흐름을 이어주는 역할을 하므로 가독성에 좋은 영향을 주는 것으로 분석이 된다. 태블릿과 인쇄물의 경우 명사구 요소(NP)가 큰 영향을 주었는데 이는 가독성이 구성과 흥미 점수를 평균화한 것이기 때문에 나타난 현상으로 간주할 수 있다. 일반적으로 명사구는 많을수록 독자에게 흥미를 줄 수 있기 때문에 양의 상관성을 크게 보이게 된다. 문장의 길이와 문장 당 단어 수가 양의 상관성을 크게 보였는데 이는 문장 내에서의 설명 등을 자세히 기술함을 의미하여 가독성을 좋게 하는 것으로 보인다. 위 결과를 통해 가독성을 결정하는 요소들이 읽기 매체마다 다른 상관관계를 가지는 것을 볼 수 있다.

4.2.2 가독성 결정 요소와 글의 구성 관점과의 관계

그림 1은 가독성 결정 요소와 가독성과의 상관관계를 글의 구성 관점 기준으로 나타낸 것이다. 읽기 매체와 상관없이 글의 길이가 길수록 글의 구성에 대한 가독성은 떨어짐을 알 수 있다. 스마트폰에서는 글의 구성 측면에서 글의 길이가 가장 큰 영향을 미쳤다. 길이 요소가 크게 영향을 주어 나머지 요소들에 대해서는 0.2 미만의 값으로 약한 상관관계를 보이게 된다. 태블릿의 경우 스마트폰과 동일하게 글의 길이가 가장 크게 영향을 주었다. 하지만 스마트폰과의 경우와는 조금 다르다. 문장의 길이, 문장 당 단어 수, 문장 당 평균 명사구 수 또한 0.2 이상의 값으로 영향을 주고 있다. 인쇄물의 경우 전체적으로 영향을 주고 있다. 위 결과는 이전 논문과 비슷한 결과를 보이고 있다[4]. 글의 구성 관점에서 글의 길이를 제외한 나머지 요소들은 읽기 매체마다 다른 특성을 보인다.

4.2.3 가독성 결정 요소와 글의 흥미 관점과의 관계

그림 2는 가독성 결정 요소와 가독성과의 상관관계를 글의 흥미 관점 기준으로 나타낸 것이다. 흥미도의 경우 읽기 매체와 상관없이 비슷한 양상의 상관관계를 보임을 알 수 있다. 어절 당 평균 어절 수의 요소가 모든 읽기 매체에 대해서 음의 상관관계를 보였다. 이는 어절 당 평균 어절 길이가 전문용어와 같은 복합명사를 나타내므로 흥미 측면에서 가독성을 떨어트릴 수 있기 때문이다. 기존 연구와 같이 문장 당 평균 명사구 및 동사구의 수에 대한 요소의 경우 명사구와 동사구의 개수가 많을수록 글의 흥미를 상승시켜주는 효과를 가져왔다. 이러한 효과는 인쇄물에서 문장의 길이와 문장 당 평균 단어 수도 같이 반영이 되어 보임을 알 수 있다.

전체적으로 스마트폰은 세 가지의 결과에 대해 글의 길이에 대해서는 공통적으로 모두 음의 상관관계를 보여 기기의 특성을 반영함을 알 수 있다. 글의 구성 측면에

서는 인쇄물은 요소들이 고루 영향이 끼침을 보이는 반면 태블릿은 문장 길이, 문장 당 단어 수, 문장 당 평균 명사구 수에 대해서 큰 영향을 받음을 보였다. 글의 흥미도 측면에서는 태블릿의 경우 인쇄물의 크기와 흡사하여 전체적으로 비슷한 양상을 보이지만 스마트폰의 경우 글의 길이나 로그 우도에서 상관관계의 부호가 달라지는 등의 차이를 보였다.

5. 결론

본 논문에서는 뉴스 기사를 이용하여 다양한 읽기 매체(인쇄물, 태블릿, 스마트폰)에서 글의 구성, 흥미를 반영한 가독성을 측정하였다. 측정된 가독성을 통해 본 논문에서는 읽기 매체에 따라 같은 글에 대해서 가독성을 결정하는 요소와 가독성이 차이가 있는지를 살펴보았다. 게다가, 글의 구성과 흥미도의 관점에서 중요하게 다뤄지는 가독성의 요소가 읽기 매체와 글의 관점마다 차이가 있는지를 살펴보았다.

실험 결과, 같은 글에 대해 읽기 매체에 따라 가독성이 달라짐을 보였다. 특히, 화면이 작은 스마트폰에서는 글의 길이와 같이 표면적인 요소에 영향을 많이 받는 것을 볼 수 있었으며, 상대적으로 화면이 큰 태블릿, 인쇄물은 구문적인 요소에 영향을 많이 받는 것을 볼 수 있었다. 또한 가독성과 글의 구성 관점에서는 기기에 따라 차이가 있음을 보였다. 글의 길이는 공통적으로 영향을 많이 받았으나, 다른 요소들은 기기마다 다르게 영향을 받음을 볼 수 있었다. 또한 가독성과 글의 흥미 관점에 대해서는 글의 구성 관점과는 달리 기기에 상관없이 일정한 관계를 가짐을 볼 수 있었다.

위의 실험 결과를 바탕으로 가독성이란 읽기 매체와 읽는 관점에 따라 다르게 반영이 되어야 함을 알 수 있다. 이에 향후 연구로는 다르게 반영된 가독성이 사용자들이 실제로 글을 읽음에 얼마나 도움이 되는지를 분석할 예정이다.

감사의 글

본 논문은 지식경제부 산업원천기술개발사업(10035348, 모바일 플랫폼 기반 계획 및 학습 인지 모델 프레임워크 기술 개발)의 지원으로 수행되었음

참고문헌

[1] J. C. Richards, J. Platt, and H. Platt, Longman dictionary of language teaching and applied linguistics, 1992.

[2] E. Dale and J. S. Chall, "The concept of readability", *Elementary English,* Vol. 26, No. 23. 1949.

[3] T. Virtanen-Ulfhielm "Linguistic Complexity in Two Major American Newspapers and The Associated Press Newswire, 1900-2000," *Master's thesis,* Abo Akademi, 2004.

[4] E. Pitler and A. Nenkova, "Revisiting Readability: A Unified Framework for Predicting Text Quality," In *Proceedings of EMNLP*, pp 186-195, 2008

[5] T. Francois and C. Fairon, "An "AI readability" formula for French as a foreign language, ", In *Proceedings of EMNLP and CoNLL*, pp. 466-477, 2012.

[6] R.Flesch, "A new readability yardstick" *Journal of Applied Psychology*, vol.32, pp.221-233, 1948.

[7] J. P. Kincaid, R. P. Fishburne, R. L. Robers, B. S. Chissom, "Derivation of New Readability Formulas(Automated Reliability Index, Fog Count and Flesch Reading Ease Formula) for Navy Enlisted Personnel," *Research Branch Report* pp.8-75, 1975.

[8] X. Yan, D. Song, and X. Li, "Conecpt-based Document Readability in Domain Specific Information Retrieval," In *Proceedings of CIKM,* pp. 540-549, 2006.

[9] R. Barzilay and M. Lapata, "Modeling Local Coherence: An Entity-based Approach," In *Proceedings of ACL*, page 141-148, 2005.