

디지털 텍스트의 음절을 이용한 운율 정보 시각화에 관한 연구

A Study on Rhythm Information Visualization Using Syllable of Digital Text

박선희, 이재중, 박진완

중앙대학교, 중앙대학교, 중앙대학교

Park seon-hee, Lee jae-joong, Park jin-wan

Chung-Ang Univ, Chung-Ang Univ, Chung-Ang Univ

요약

정보화 시대가 빠르게 성장하면서 디지털 텍스트의 양도 증가하고 있다. 이에 따라 수많은 디지털 텍스트를 파악하기 위한 시각화 사례가 증가하고 있다. 기존의 디지털 텍스트 시각화 디자인은 스템밍 알고리즘(stemming algorithm)의 도입과 단어 빈도수를 추출하여 주제를 형성화하여 텍스트의 의미를 부각시키고 문장과 문장을 연결해주는 것에 치중하고 있다. 이에 디지털 텍스트의 정서적인 느낌을 시각화할 수 있는 운율을 표현하는 것에 있어서 미흡했던 부분이 사실이다. 운율을 보다 효과적으로 표현할 수 있는 음운단위로는 음절을 들 수 있다. 문장에서 음절은 단어나 구, 문장의 발음에 가장 기본적인 발음 단위가 된다. 이를 기본으로 강세, 성조, 운율 요소들의 길이 등이 음절에 기반을 두고 있다. 음절을 정의하는 것과 가장 밀접한 연관이 있는 공명도(sonority)는 발화할 때 폐의 공기 흐름과 운동 에너지(Kinetic energy)를 공명도로 명시되는 음향에너지(acoustic energy)로 표현한 것이다. 본 연구는 이러한 관점에서 디지털 텍스트의 속성인 음절을 기반으로 음운론적 정의와 특성을 살펴보고 운율을 다이어그램을 통해 시각화하기 위한 방법을 연구한다. 실험을 통해 디지털 텍스트를 발음기호로 변환한 후, 모든 언어속의 리듬에서 출발된 음절의 공명도를 사용하고 디지털 텍스트를 음절화하여 운율 정보를 이미지로 시각화한다. 운율 정보를 시각화함으로써 디지털 텍스트의 음절 정보를 알 수 있고, 디지털 텍스트의 정서를 다이어그램을 통해 체계적인 공식에 의하여 사용자의 이해를 돕도록 표현한다. 이에 해당 텍스트의 운율을 보다 쉽게 파악하도록 설계하여 디지털 정보 시각화를 구현하는데 그 목적을 두고 있다.

Abstract

As the information age grows rapidly, the amount of digital texts has been increasing as well. It has brought an increasing of visualization case in order to figure out lots of digital texts. Existing visualized design of digital text is merely concentrating on figuration of subject word through adoption of stemming algorithm and word frequency extraction, prominence of meaning of text, and connection in between sentences. So it is a fact that expression of rhythm that can visualize sentimental feeling of digital text was insufficient. Syllable is a phoneme unit that can express rhythm more efficiently. In sentences, syllable is a most basic pronunciation unit in pronouncing word, phase and sentence. On this basis, accent, intonation, length of rhythm factor and others are based on syllable. Sonority, which is most closely associated with definitions of syllable, is expressed through air flow of igniting lung and acoustic energy that is specified kinetic energy into sonority. Seen from this perspective, this study examines phonologic definition and characteristics based on syllable, which is properties of digital text, and research the way to visualize rhythm through diagram. After converting digital text into phonetic symbol by the experiment, rhythm information are visualized into images using degree of resonance, which was started from rhythm in all languages, and using syllable establishment of digital text. By visualizing syllable information, it provides syllable information of digital text and express sentiment of digital text through diagram to assist user's understanding by systematic formula. Therefore, this study is aimed at planning for easy understanding of text's rhythm and realizing visualization of digital text.

I. 서론

1. 연구 배경 및 목적

사용자들이 받아들이는 정보의 양이 많아진다고 해서 반드시 더 좋은 판단을 할 수 있는 것은 아니라는 사실이 점점 명확해지고 있다. 무분별하게 넘쳐나는 정보 앞에서 사용자가 필요로 하는 요소를 찾기란 힘든 실정이다. 그에 따라 수많은 정보를 내포하고 있는 디지털 텍스트를 파악하기 위한 시각화 사례가 증가하고 있다. 기존의 디지털 텍스트 시각화 디자인에서는 스타밍 알고리즘과 단어빈도수를 추출하는 과정을 통해 디지털 텍스트의 내용을 함축하는 주제어를 형상화하고 관련있는 단어가 포함된 문장과 문장을 연결하는 것에 치중하고 있다. 이에 본 연구에서는 디지털 텍스트의 운율 정보를 시각화하고자 음운기술의 최적 단위인 음절을 하나의 정보로 간주하고 공명도 이론을 이용하여 운율을 시각화한다. 또한 반복되는 각운을 추출하여 해당 디지털 텍스트의 정서적인 느낌을 표현하고자 한다.

II. 음절과 공명도 이론

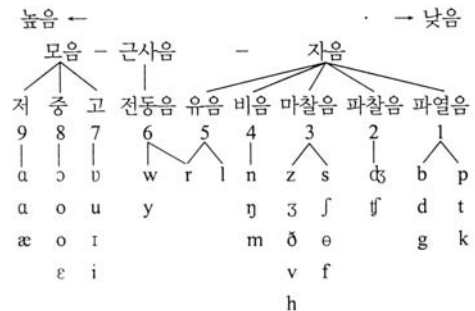
1. 음절의 정의

음절은 현재 중요한 음운단위로 널리 인정받고 있다. 음절은 음운단위의 기본으로 강세, 성조, 운율 요소들의 길이 등이 음절에 그 기반을 둔다. 모든 음절은 분절음이 자리하는 최소한의 단위이기도 하다. 분절음들은 발음되기 위해서 음절이 되고, 음절은 단어나 구, 문장의 발음에 가장 기본적인 발음 단위가 된다[1]. 일찍부터 운문은 음절을 바탕으로 쓰였으며, 음절은 음성 현상의 설명을 쉽고 간결하게 해주는 기능을 한다. 이러한 점에서 음절과 밀접한 연관이 있는 공명도(sonority)에 대해 알아보기로 한다.

2. 공명도의 정의

분절음은 하나의 단위로서 각각 자신의 공명도를 갖는다. 각 분절음의 공명도가 모여 하나의 특징적인 단위를 형성하는데 그것이 바로 음절이다. 음절은 모음을

중심으로 분절음들의 올림 정도 즉, 공명도(Sonority degree)가 좌우 방향으로 낮게 표현된다. 이에 분절음의 공명도는 음절 형성의 기준이 된다. 공명도란 어떤 소리가 일정한 에너지를 사용하여 내는 소리의 크기로 할 수 있다. 음향학적으로 공명이 크게 일어나거나 에너지가 큰 것을 공명도가 높은 것이라고 하고, 지각적인(perceptual)측면으로는 멀리까지 크게 잘 들리는 것이 공명도가 높은 것이라고 할 수 있다. 영어의 공명도 층위(Sonority Hierarchy)는 다음과 같다.



▶▶ 그림 1. 영어 분절음의 공명도 층위

높은 값을 갖는 공명도를 [강]으로 표시하고, 낮은 값을 갖는 공명도를 [약]으로 표시하면 음절의 수를 확인할 수 있다. 또한 공명도가 높은 음절 꼭대기의 정점(peak point)을 기준으로 분절음들이 가진 공명도의 변화가 음절 내의 운율성(rhythmicity)을 확인할 수 있다 [2]. 이에 음절 내의 분절음이 이루는 [강]과 [약]의 교차로 음절 내의 운율성이 나타난다. 이와 같은 음절내의 분절음이 이루는 운율성과 강세와 약세의 음절 배분에 의한 단어의 운율성이 운율을 형성한다. 또한 어구의 연결에 의하여 나타나는 모든 구절의 운율로 연결되어 발화의 운율이 형성된다.

3. 텍스트의 운율성

텍스트 내에서의 강세는 서로 조화를 이루며 일정한 시간 간격을 유지하며 나타날 때 운율감이 나타난다. 자연스러운 운율을 만들려면 시간 개념이 들어가야 한다. 운율은 주로 공명성을 이루는 모음류(vocalic)의 단음화(shortening)와 장음화(lengthening)와 축약, 탈락, 삼입에 의한 현상으로 나타나고 자음의 변화와 탈락과 삼입으로 나타난다. 또한 발화로서의 운(rime)과 울(meter)을 가지고 휴지(pause)와 함께 리듬을 형성하여

의미를 전달한다.

Ⅲ. 시각화를 위한 사전 연구

1. 정보시각화 소개

1.1. 정보의 정의

우리는 항상 정보에 둘러싸여 살아가고 있다. 모든 정보는 그 목적에 따라 의미와 가치를 지니며 생산자와 사용자의 관점에 따라 다양하게 정의된다. 정보가 되기 전 단계인 데이터는 정보로서 완전한 형태와 의미를 갖추지 못한 모습을 가르킨다. 정보는 가공되지 않은 데이터에 의미와 가치를 부여하여 조직화하고 변화시키며 그 형태와 형식을 지닌다.

1.2. 정보 시각화의 정의

한 연구에 따르면 인간의 전체 감각 중에서 시각이 80% 이상을 차지한다고 한다. 언어나 수치로만 된 데이터와 정보를 마음속에서 그림이나 이미지와 같은 형태로 바꾸어 사고하는 성향이 있다. 이것을 우리는 흔히 시각화라 한다. 시각화는 마음속의 이미지(心象)를 만드는 것, 또는 실제 보는 것과 같이 상상하고 기억하는 것이라는 사전적 의미가 있으며[3], 인간의 인지 활동과 깊은 관련을 맺고 있다. 정보 시각화란 사용자에게 더 효율적으로 정보를 전달하기 위하여 그래픽 요소를 활용하여 데이터가 정보로서 의미를 생성되도록 형상화하는 것을 뜻한다[4]. 따라서 정보의 시각화는 시각화되지 않은 정보를 받아들일 때보다 인지 능력을 증폭시킬 수 있다.

1.3. 정보 시각화의 효과

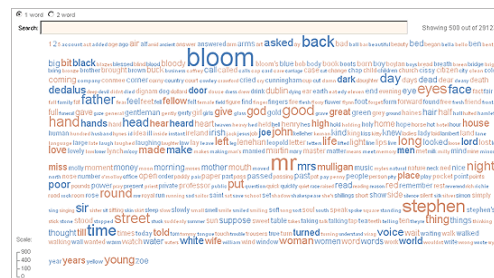
시각화된 정보는 시각화되지 않은 정보에 비해 정보를 직관적으로 이해하기 편리하고, 많은 데이터를 동시에 차별적으로 보여줄 수 있다. 또한 눈에 보이지 않는 원자의 구조, 전자의 흐름 등과 같은 난해한 과학 원리의 지각적 추론을 가능하게 한다. 그리고 정보에 감성을 부여함에 있어서 보는 이의 흥미를 유발하고 표현에 따라 웃음, 슬픔 등의 감정 표현이 가능하게 하여 이

를 통해 인간의 경험을 풍부하게 한다. 또한 일반적으로 사람들은 그림이 문자보다 쉽다고 생각하는 경향이 있다. 시각화된 정보를 더 친근하게 느낀다. 더 나아가 더 넓은 계층의 사람들에게 쉽게 다가갈 수 있다. 이에 시각화의 궁극적인 목표인 정보 전달의 효율성 증대에 초점을 맞추어야 하며 모든 사람들에게 쉽게 다가가기 위해 프로세스에 가장 신경을 써야 한다.

2. 텍스트 시각화의 사례

2.1. Tag Cloud

Tag Cloud는 주어진 디지털 텍스트에서 단어의 빈도수를 시각화한다. Tag Cloud는 빈도수가 높은 주제를 눈에 잘 들어오도록 각각의 단어들의 빈도수를 계산하고 빈도수의 양에 맞게 텍스트 크기를 표현하여 가시화를 높이고자 하였으며, 발음기호를 제거한 후, 기존에 널리 알려진 스테밍 알고리즘 기법을 적용하여 'the'와 같은 조사를 제거하였다. 후에, Tag Cloud는 이미지가 아닌 텍스트를 나열하는 방식으로 텍스트의 크기만으로 시각화하여 보여주고 있다.



▶▶ 그림 2. Tag Cloud 구현 결과

2.2. What Colors Tell About a Text

What Colors Tell About a Text는 품사 별로 색을 지정해 주어 문서의 정서를 시각화한다. 색상으로 지정된 단어들은 디지털 텍스트의 장르와 문장 구조, 문체를 상세하게 제공한다. 소설과 과학 문서를 비교한 경우, 소설 문서가 과학 문서보다 더 밝은 색상 패턴으로 구성되어 나타난 것을 알 수 있다. 그러므로 문서 장르를 분석하는데 용이하다. 또한 이러한 색상 패턴은 디지털 텍스트에 숨겨진 소리와 멜로디와 관련이 깊다고 주장하고 있다.



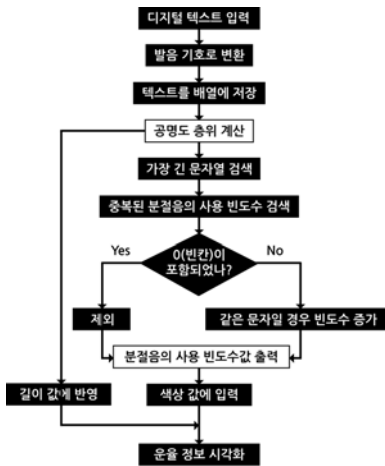
▶▶ 그림 3. Whats Color Tell About a Text 구현 결과

IV. 디지털 텍스트를 이용한 시각화 구현

1. 디자인 설계

본 실험을 통해 디지털 텍스트의 운율 정보 시각화를 구현하여 사용자가 텍스트를 직접 읽지 않고도 디지털 텍스트 운율을 이해하는데 용이하도록 한다. 본 실험에서는 크게 디지털 텍스트 입력, 공명도 이론을 제한한다. 먼저 디지털 텍스트를 입력하여 발음기호로 변환하여 추출한 후, 공명도에 의한 수치 값을 이용하여 각 분절음들의 수치 값을 출력한다. 그 후, 얻은 수치값을 이용하여 디지털 텍스트에서 사용된 각 분절음들의 빈도수를 찾아내어 그에 따른 운율을 시각화하고자 한다. 진행 과정에서는 자바 기반 언어인 프로세싱(processing)을 사용하였다.

2. 디지털 텍스트 시각화 순서도

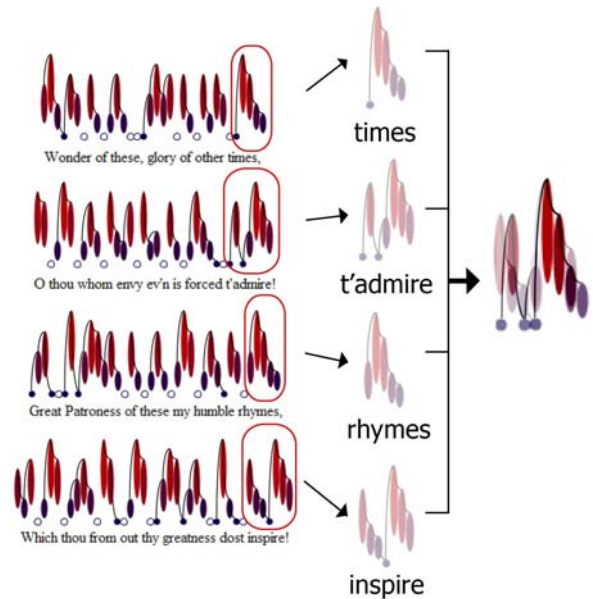


▶▶ 그림 4. 디자인 제작 순서도

3. 디자인 과정 및 결과

3.1. 공명도 층위 적용 실험

먼저, 본 실험에 앞서 공명도 층위를 검사하여 해당 영어 발음을 9가지 단계로 나누어 운율의 강약을 가시화시켜 본다. 디지털 텍스트는 서정시 “THE LADY MARY” 를 입력하였다. 각각의 분절음을 시각화하여 나타내면 [그림5]와 같다. 탐색에서 붉은 계열로 갈수록 공명도의 값은 증가하며, 그에 따라 원의 세로 반지름 값도 증가한다. 한 단어에는 정점을 이루는 분절음 하나 이상 존재하며, 그 분절음을 중심으로 그 앞이나 뒤로 점차 낮아지는 공명도의 값을 갖는 분절음들이 온다는 것을 확인할 수 있다. 또한 공명도 층위를 시각화했을 때 반복되는 각운의 공명도 층위 단계가 일관성 있게 공통된다는 것을 확인할 수 있다.

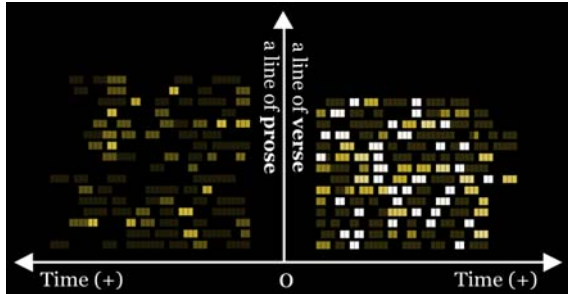


▶▶ 그림 5. 공명도 층위 시각화 구현 결과

3.2 반복되는 운율의 빈도수 확인 실험

공명도 층위에 해당하는 9단계의 숫자데이터를 출력한 후에 그 중 가장 긴 문자열을 검색하여 가장 긴 문자열부터 문자열이 최소 2개가 될 때까지 중복되는 분절음을 찾는다. 중복된 분절음들을 찾은 후 그 위치에 해당 빈도수를 입력하여 해당 분절음의 빈도수를 확인할 수 있었다. 이렇게 계산한 빈도수 값을 가지고 디지털 텍스트에서 일정한 운율을 갖고 있는지를 알아보기

위해 운문과 산문을 가지고 각각의 중복된 운율의 빈도수를 시각화하고 총 문자열에서의 중복된 운율의 백분율을 출력해 보았다.

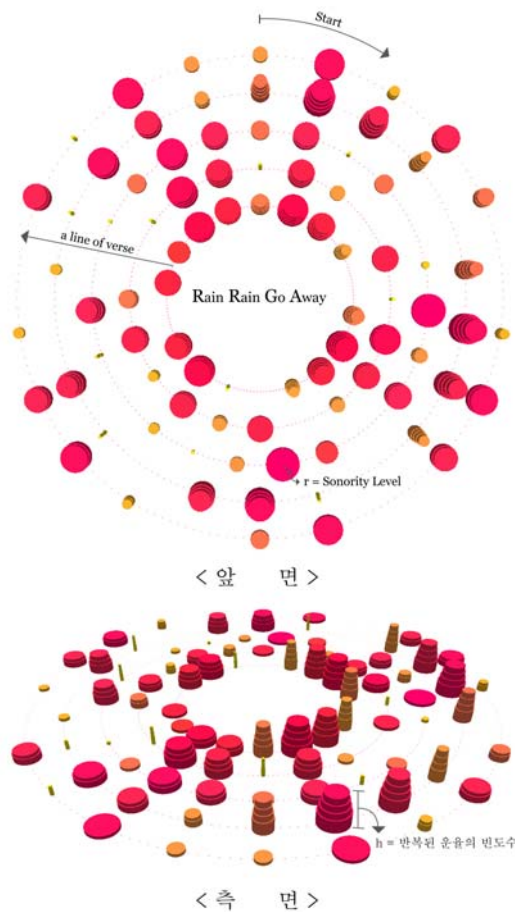


▶▶ 그림 6. 반복되는 운율 빈도수 구현 결과

[그림 6에서 왼쪽(산문)과 오른쪽(운문) 이미지를 비교해 보았을 경우 오른쪽 이미지를 나타내는 운문이 산문보다 전체적으로 밝은 색상의 사각형 패턴이 많이 나타나 있는 것을 볼 수 있다. 실험 결과, 운문은 총 문자열에서 64.17%의 반복된 운율을 가진 분절음이 나타나며, 산문에서는 27.59%정도가 나타나고 있다는 것을 확인할 수 있었다.

3.3. 디자인 구현 결과

본 실험에서는 앞에서 구현된 공명도 층위를 이용한 실험과 그에 따른 반복되는 운율의 빈도수를 가지고 3D 환경을 적용하여 실험 해보았다. 공명도 층위 그래프의 높낮이를 통해 운율의 강약을 표현하기 위해 분절음들을 각각의 원의 크기를 표현한다. 또한 공명도 층위와 문장의 길이 값에 색상 값에 적용한다. 문장이 길어질수록 녹색 계열을 나타내고, 짧은 문장일수록 붉은 계열이 나타난다. 원들의 운율을 비교하기 위해 디지털 텍스트의 한 문장 단위로 나누어 문장 길이에 따라 적절한 위치를 찾아내어 12시 방향에서부터 360도 회전을 이룬다. 반복되는 운율의 빈도수는 측면에서 보았을 때 하나의 다이어그램을 형성하여 시각화하도록 한다. [그림7에서 상단에 시작되는 위치를 중심으로 한 분절음들의 원의 크기를 통해 운율이 비슷하다는 것을 확인할 수 있다. 이에 반해, [그림9]에서는 문장이 많아짐에 따라 분절음에 해당하는 원의 크기 또한 작게 시각화되어 공명도 층위를 알기 어렵다는 것을 확인할 수 있다.



▶▶ 그림 7. 운문을 입력했을 경우 최종 디자인 구현 결과



▶▶ 그림 8. 산문을 입력했을 경우 최종 디자인 구현 결과

V. 결론 및 향후 연구 방향

본 실험은 디지털 텍스트의 운율 정보를 시각화하는데 관련이 깊은 공명도 층위와 반복되는 운율의 빈도수를 추출하여 최종적으로 운율 정보시각화 결과물을 도출해내는 형태이다. 본 실험에서 운문과 산문을 비교한 경우, 운문이 산문보다 더 따뜻한 색감의 패턴으로 구성된 것을 알 수 있었다. 결과적으로 디지털 텍스트의 운율 정보 시각화는 문서 장르를 분석함에 있어서 용이하며, 디지털 텍스트를 직접 읽었을 때 발견하지 못한 운율감을 시각화하여 문서의 감정을 전달함에 있어서 유용하다. 향후에는 본 연구에서 구현한 분절을 빈도수를 더 심도있게 분석하여, 더 정확한 결과물을 도출하고자 한다.

■ 참고 문헌 ■

- [1] 김인옥 “음절이론과 음운화과정의 분석”, 금오공과대학교 석사학위논문, 2005
- [2] 김기섭 “영어의 재구조화와 발음 변화”, 한국문화사, p.28~363, 2005
- [3] Robert Spence "Information Visualization", Addison - Wesley, p.1, 2001
- [4] 오병근 "정보디자인교과서", 안그래픽스, p.19~101, 2008
- [5] 이유진 “시각적 텍스트로 표현한 한국 현대시(詩)의 시각은유 연구”, 정보디자인학연구, 2008
- [6] 전상범 “영어음성학개론”, 을유문화사, 2006

본 연구는 문화체육관광부 및 한국문화콘텐츠진흥원의 2008문화기술(CT) 및 공연예술기술개발지원사업과 2008년도 2단계 두뇌한국(BK)21 사업의 연구결과로 수행되었음

■ 부 록 ■

1) Rain Rain go Away

Rain rain go away,
Come again another day.
Little Johnny wants to play;

Rain, rain, go to Spain,
Never show your face again!

2) THE OUTLINE OF SCIENCE

Evolution, as we have seen in a previous chapter, is another word for race history. It means the ceaseless process of Becoming, linking generation to generation of living creatures. The Doctrine of Evolution states the fact that the present is the child of the past and the parent of the future. It comes to this, that the living plants and animals we know are descended from ancestors on the whole simpler, and these from others likewise simpler, and so on, back and back till we reach the first living creatures, of which, unfortunately, we know nothing. Evolution is a process of racial change in a definite direction, whereby new forms arise, take root, and flourish, alongside of or in the place of their ancestors, which were in most cases rather simpler in structure and behaviour.

The rock record, which cannot be wrong, though we may read it wrongly, shows clearly that there was once a time in the history of the Earth when the only backboned animals were Fishes. Ages passed, and there evolved Amphibians, with fingers and toes, scrambling on to dry land. Ages passed, and there evolved Reptiles, in bewildering profusion. There were fish lizards and sea serpents, terrestrial dragons and flying dragons, a prolific and varied stock. From the terrestrial Dinosaurs it seems that Birds and Mammals arose. In succeeding ages there evolved all the variety of Birds and all the variety of Mammals. The question is whether similar processes of evolution are still going on. We are so keenly aware of rapid changes in mankind, though these concern the social heritage much more than the flesh and blood natural inheritance, that we find no difficulty in the idea that evolution is going on in

mankind. We know the contrast between modern man and primitive man, and we are convinced that in the past, at least, progress has been a reality. That degeneration may set in is an awful possibility involution rather than evolution but even if going back became for a time the rule, we cannot give up the hope that the race would recover itself and begin afresh to go forward.