



(19) 대한민국특허청(KR)  
(12) 공개특허공보(A)

(11) 공개번호 10-2015-0047259

(43) 공개일자 2015년05월04일

(51) 국제특허분류(Int. Cl.)

G06F 17/00 (2006.01)

(21) 출원번호 10-2013-0127112

(22) 출원일자 2013년10월24일

심사청구일자 2013년10월24일

(71) 출원인

서강대학교산학협력단

서울특별시 마포구 백범로 35 (신수동, 서강대학교)

(72) 발명자

김주섭

서울 양천구 목동남로2길 60-30, 104동 401호 (신정동, 세양청마루아파트)

웨인 드 프레머리

경기 고양시 덕양구 푸른마일로 16, 201동 1905호 (고양동, 푸른마을2단지)

(74) 대리인

김일환

전체 청구항 수 : 총 9 항

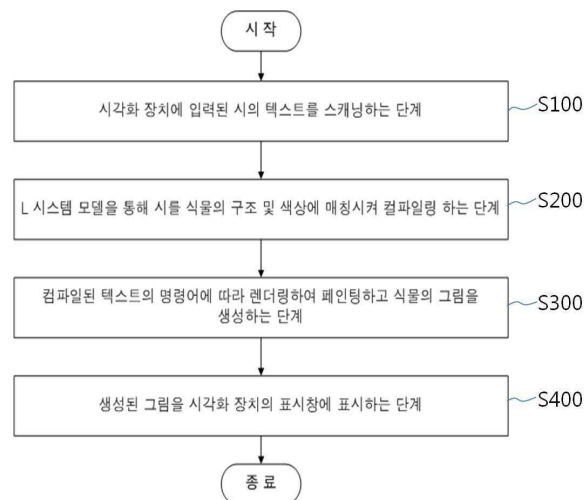
(54) 발명의 명칭 L 시스템을 이용한 시 텍스트 시각화 방법

### (57) 요약

본 발명은 시 텍스트 시각화 방법에 관한 것으로, (a) 시각화 장치에 입력된 시의 텍스트를 순서대로 스캐닝하는 단계; (b) L 시스템 모델을 통해 상기 시의 구조를 식물의 구조와 매칭시키고, 상기 텍스트의 형태소 및 모음을 식물의 채도와 명도와 매칭시켜 컴파일링 하는 단계; (c) 컴파일된 상기 텍스트의 명령어에 따라 렌더링하여 페인팅하고, 상기 식물의 그림을 생성하는 단계; 및 (d) 생성된 상기 그림을 상기 시각화 장치의 표시창에 표시하는 단계를 포함한다.

이와 같이 본 발명은 시 텍스트의 시각화에 정보 전달 기능 외에 심미성을 부가하여 시각화 결과 자체가 엔터테인먼트적인 요소를 가지고 가며 직관적인 영감까지 줄 수 있는 창작물이 되도록 하는 새로운 시각화 방법을 제공한다.

대표도 - 도6



## 명세서

### 청구범위

#### 청구항 1

- (a) 시각화 장치에 입력된 시의 텍스트를 순서대로 스캐닝하는 단계;
- (b) L 시스템 모델을 통해 상기 시의 구조를 식물의 구조와 매칭시키고, 상기 텍스트의 형태소 및 모음을 식물의 채도와 명도와 매칭시켜 컴파일링 하는 단계;
- (c) 컴파일된 상기 텍스트의 명령어에 따라 렌더링하여 페인팅하고, 상기 식물의 그림을 생성하는 단계; 및
- (d) 생성된 상기 그림을 상기 시각화 장치의 표시창에 표시하는 단계를 포함하는 것을 특징으로 하는 L 시스템 모델링을 이용한 시 텍스트의 시각화 방법.

#### 청구항 2

제1항에 있어서,

상기 시각화 장치는 PC, 태블릿 PC, 휴대폰, PDA 중 어느 하나인 것을 특징으로 하는 시 텍스트의 시각화 방법.

#### 청구항 3

제1항 또는 제2항에 있어서,

상기 (b) 단계는,

형태소 분석기를 이용하여 상기 텍스트의 형태소를 분석하여 구분하는 단계를 포함하는 것을 특징으로 하는 시 텍스트의 시각화 방법.

#### 청구항 4

제3항에 있어서,

상기 (b) 단계는,

상기 시의 구조를 어절, 행 및 연으로 구분하고, 상기 식물의 잎사귀, 가지및 몸통의 구조로 매칭하는 것을 특징으로 하는 시 텍스트의 시각화 방법.

#### 청구항 5

제4항에 있어서,

상기 (b) 단계는

상기 형태소의 품사에 따라 채도를 매칭시키고, 상기 텍스트의 모음에 따라 명도를 매칭시키는 것을 특징으로 하는 시 텍스트의 시각화 방법.

#### 청구항 6

제5항에 있어서,

상기 모든 잎사귀의 색을 녹색으로 설정하는 것을 특징으로 하는 시 텍스트의 시각화 방법.

#### 청구항 7

제5항에 있어서,

상기 (b) 단계는,

상기 텍스트의 글자수와 비례하여 상기 식물의 잎사귀 두께를 매칭시키는 것을 특징으로 하는 시 텍스트의 시각화 방법.

#### 청구항 8

제5항에 있어서,

상기 어절이 반복되는 수를 나타내는 중복도와 비례하여 상기 몸통의 두께를 매칭시키는 것을 특징으로 하는 시 텍스트의 시각화 방법.

#### 청구항 9

제7항에 있어서,

상기 중복도는,

$$S = \frac{1}{n} \sum S_j$$

(n = 한 시에 포함된 어절의 개수)

$$S_j = \frac{1}{n} \sum_{i \neq j} dist(어절_i, 어절_j), 0 \leq dist(.) \leq 1$$

n = 한 시에 포함된 어절의 개수 - 1

dist(어절<sub>i</sub>, 어절<sub>j</sub>) = 서로 다른 어절간 유사도

(S<sub>j</sub>는 j당 중복도를 나타내고, n은 한 시에 포함된 어절의 개수이다.)

와 같은 식으로 계산되는 것을 특징으로 하는 시 텍스트의 시각화 방법.

### 발명의 설명

#### 기술 분야

[0001]

본 발명은 텍스트 시각화 방법에 관한 것으로, 보다 상세하게는 기능성 및 심미성을 갖추어 용이하게 시각화할 수 있는 L 시스템을 이용한 시 텍스트 시각화 방법에 관한 것이다.

#### 배경 기술

[0002]

미국 시인 Emily Dickinson(1830-1886)는 종종 시를 거꾸로 읽을 때 마음을 사로 잡는 무언가를 발견한다고 하였다. 이것을 두고 비평가 Jerome McGann은 시를 거꾸로 읽는 것은 텍스트가 주는 느낌을 해체하여 감상하게 도와주는 매우 좋은 방법으로 독자를 다른 방법으로는 알기 불가능할 수 있는 것들을 상상할 수 있는 비평적인 위치로 인도한다고 하였다. Dickinson와 McGann의 연장선상에서 Stephen Ramsay는 텍스트를 이해하는 최선의 한

방법은 텍스트를 변형하여 여러 다른 시각에서 보면서 그러한 변형이 어떠한 영향을 미치는지 알아보는 것이라고 하였다. Ramsay는 텍스트를 비평적으로 유용한 방향으로 변형(deformance)하는데 컴퓨터 테크놀로지가 유용할 수 있다고 주장한다. 그는 이러한 과정을 Algorithmic Criticism이라 부른다.

[0003] 컴퓨터를 사용하여 텍스트를 새로운 관점에서 바라보는 주된 방법은 시각화(visualization)라고 할 수 있다. 현재 거의 대부분의 텍스트 시각화에서 쓰이는 시각적 도구는 인포그래픽스(Infographics)에서 많이 쓰이는 chart, bar graph, word map 등이다. 이러한 방법은 정보 전달 면에서 명확하고 정확하다.

[0004] 텍스트 데이터는 문학작품뿐만 아니라 뉴스, 잡지, 웹, 성경 등 텍스트로 이루어진 모든 데이터를 일컫는다. 텍스트 데이터는 기본적으로 형식과 내용으로 구별 지어 볼 수 있는데 각각이 텍스트 데이터의 시각화 대상이 된다. 형식은 문단, 문장, 어절, 형태소, 음절, 음운 등 형태론적 특성을 가리키며 내용은 반복되는 키워드나 단어의 긍정적/부정적 의미 등을 포함한다.

[0005] 텍스트 데이터 시각화는 그 목적에 따라 그 방법을 나눌 수 있다. 우선 정보의 직관적 전달이 목적인 경우가 가장 일반적이라고 할 수 있는데 Stefanie Posavec과 Greg McInerny는 'The Evolution of The Origin of Species' 에서 영국 생물학자 찰스 다윈의 저서인 종의 기원 1판부터 6판까지를 시각화 하였다(도 1).

[0006] 각 문단은 하나의 선에서 나온 하나의 잎으로 표현되어 지고 잎 안에 그려진 웨지(wedge)모양 하나하나를 한 문장을 의미하고 색으로 인코딩 되어 있다. 그리고 각 색은 그 문장이 다음 버전의 책에서 계속 존재하는지 사라지는지 가리킨다. 이렇게 함으로써 직관적으로 책이 개정되면서 어떻게 변화되었는지 알 수 있게 도와준다.

[0007] 이러한 기능성 말고 시각화는 심미성을 목적으로 수행되기도 한다. 즉, 또 다른 시각 예술 작품의 재료로 텍스트가 이용되어 지는 것이다. Boris M의 2003년 (도 2), 2006년(도 3) 독일 국제 문학 페스티벌 포스터는 그 해 출판된 시들을 모아 놓고 알고리즘을 적용하여 컴퓨터가 자동으로 그린 그림을 보여 주고 있다. 그의 2003년 포스터는 각 알파벳이 선의 이동 방향을 바꾸어 그리게 하는 명령어로 작동하게 함으로써 시 하나를 하나의 컴퓨터 프로그래밍처럼 작동하도록 바꾸었다.

[0008] 또한 2006년 작품에서는 각 알파벳에 숫자를 할당하여 한 단어를 그 단어를 이루는 숫자들의 합으로 치환하여 그 값에 비례하는 크기를 가지는 빨간 원으로 표시한 후, 이러한 단어들을 시의 길이에 비례하는 원주위에 배치하고 반복되어 사용되어지는 단어를 회색 선으로 연결하여 표현함으로써 마치 물질의 원자 내부를 보는 것처럼 형상화하였다. 이러한 것은 정보 제공보다는 새로운 창작을 하는데 시각화가 쓰인 예제이다.

[0009] 기능성과 심미성을 동시에 가지고 가기도 하는데 Stefanie Posavec은 그의 프로젝트 'writing without words' 에서 문학 작품을 문단, 문장, 단어 등으로 분리하여 계층적으로 선으로 연결된 꽃잎의 형태로 표현하였다(도 4). 또한 작품의 주제에 따라 11가지의 색상을 할당함으로써 작품의 형식과 내용을 동시에 표현하려고 하였다. 비록 약하지만 어느 정도 정보 전달의 기능성도 있으면서 아름답게 시각적 작품으로 재창작 된 예제라고 할 수 있다.

[0010] 이밖에도 관련 국내 연구로는 박진완의 'Visual Genealogy' (도 5)를 들 수 있는 데 이 작품에서는 한국의 족보를 시각화하여 나타냄으로써 기능성이나 심미성 보다는 그 안에서 새로운 이야기를 창출해내는 용도로 시각화가 쓰인 예제이다.

[0011] 그러나 종래의 텍스트의 시각화 방법은 정보전달을 위한 기능성이나 새로운 창작을 위한 심미성을 구분하여 그 목적에 적합한 방법을 제공하거나, 기능성 및 심미성을 동시에 만족하는 시각화 방법을 제안하고 있지만 그 전달 효과가 떨어지고, 대중적 접근성이 떨어진다는 문제점이 있다.

## 선행기술문헌

### 특허문헌

- [0012] (특허문헌 0001) 대한민국 등록특허공보 제10-0359265호(2002년10월18일)  
(특허문헌 0002) 대한민국 공개특허공보 제10-2011-0070386호(2011년06월24일)  
(특허문헌 0003) 대한민국 등록특허공보 제10-1055513호(2011년08월02일)

## 발명의 내용

### 해결하려는 과제

- [0013] 상술한 문제를 해결하고자 하는 본 발명의 과제는 시라는 텍스트 시각화하는 새로운 대안적 표현 방법을 제공하고, 종래의 텍스트를 비평적으로 읽는 방법(close reading 혹은 distant reading등)을 대체하려는 것이 아니라, 종래의 방법을 도와 텍스트를 변형/해체하여 새로운 각도에서 조망할 수 있게 도와주는 새로운 방법의 시 텍스트 시각화 방법을 제공하고자 함이다.
- [0014] 또한, 시 텍스트에 대한 시각적 정보를 제공함과 동시에 심미적인 기능을 제공하는 시각화 방법을 제공하고자 하는 것으로, 시각화 결과물이 단순히 분석 결과를 보여주는 인포그래픽스가 아닌 또 하나의 창작물이 될 수 있는 시각화 방법을 제공하고자 함이다.

### 과제의 해결 수단

- [0015] 상술한 과제를 해결하고자 하는 본 발명의 특징은 (a) 시각화 장치에 입력된 시의 텍스트를 순서대로 스캐닝하는 단계; (b) L 시스템 모델을 통해 상기 시의 구조를 식물의 구조와 매칭시키고, 상기 텍스트의 형태소 및 모음을 식물의 채도와 명도와 매칭시켜 컴파일링 하는 단계; (c) 컴파일된 상기 텍스트의 명령어에 따라 렌더링하여 페인팅하고, 상기 식물의 그림을 생성하는 단계; 및 (d) 생성된 상기 그림을 상기 시각화 장치의 표시창에 표시하는 단계를 포함한다.
- [0016] 여기서, 상기 시각화 장치는 PC, 태블릿 PC, 휴대폰, PDA 중 어느 하나인 것이 바람직하고, 상기 (b) 단계는, 형태소 분석기를 이용하여 상기 텍스트의 형태소를 분석하여 구분하는 단계를 포함하는 것이 바람직하다.
- [0017] 또한, 바람직하게는 상기 (b) 단계는, 상기 시의 구조를 어절, 행 및 연으로 구분하고, 상기 식물의 잎사귀, 가지 및 몸통의 구조로 매칭하는 것일 수 있고, 상기 (b) 단계는, 상기 형태소의 품사에 따라 채도를 매칭시키고, 상기 텍스트의 모음에 따라 명도를 매칭시키는 것일 수 있다.
- [0018] 더하여, 상기 모든 잎사귀의 색을 녹색으로 설정하는 것이 바람직하고, 상기 (b) 단계는, 상기 텍스트의 글자수와 비례하여 상기 식물의 잎사귀 두께를 매칭시키는 것이 바람직하며, 상기 어절이 반복되는 수를 나타내는 중복도와 비례하여 상기 몸통의 두께를 매칭시키는 것이 바람직하다.
- [0019] 그리고, 상기 중복도는,

$$S = \frac{1}{n} \sum S_j$$

(n = 한 시에 포함된 어절의 개수)

$$S_j = \frac{1}{n} \sum_{i \neq j} \text{dist}(\text{어절}_i, \text{어절}_j), 0 \leq \text{dist}(\cdot) \leq 1$$

n = 한 시에 포함된 어절의 개수 - 1  
 $\text{dist}(\text{어절}_i, \text{어절}_j)$  = 서로 다른 어절간 유사도

- [0021]
- [0022] ( $S_j$ 는 j당 중복도를 나타내고, n은 한 시에 포함된 어절의 개수이다.)
- [0023] 와 같은 식으로 계산되는 것이 바람직하다.

### 발명의 효과

- [0024] 이와 같이 본 발명은 텍스트 중 '시' 라는 문학 장르에 국한하여 알고리즘 기반 모델링(procedural modeling)을 사용하여 각각의 시를 그 고유의 특성을 반영하는 자연의 유기적인 나무 형태로 디지털 공간에서 재창조해 내는 방법을 제공한다. 즉, 시 한편은 그 시의 특징들을 은유적으로 표현하는 한 그루의 나무로 컴퓨터 알고리즘을 이용해 모습이 변형되어 재탄생되어 지게 함으로써, 시각화에 정보 전달 기능 외에 심미성을 부가하여 시각화 결과 자체가 엔터테인먼트 적인 요소를 가지고 가며 직관적인 영감까지 줄 수 있는 창작물이 되도록 하는 새로운 시각화 방법을 제공한다.

[0025] 또한 본 발명은 텍스트의 비평적 감상을 위해 텍스트를 변형하는데 사용되는 디지털 공간상의 시각적 도구가 심미적/엔터테인먼트적 요소를 가지고 간다는 점이 종래의 시각화 방법과 차별화되고, 시각적으로 변형된 시가 새로운 심미적 감정을 불러일으키고, 원문을 단순히 분석하는 것이 아니라 새로운 시각으로 감상할 수 있게 한다는 장점이 있고, 컴퓨터라는 시각화 장치를 이용하여 정확한 정보 전달을 뛰어 넘어 심미성까지 만들어 낼 수 있는 방법을 제공한다.

### 도면의 간단한 설명

[0026] 도 1은 찰스 다윈의 '종의 기원'을 종래의 방법으로 시각화한 결과를 나타내고,  
 도 2는 보리스 풀러의 2003년 독일 국제문학 페스티벌에 출품한 시를 컴퓨터 알고리즘을 이용해 시각화한 포스트를 나타내고,  
 도 3은 보리스 풀러의 2006년 독일 국제문학 페스티벌에 출품한 시를 컴퓨터 알고리즘을 이용해 시각화한 포스트를 나타내고,  
 도 4는 문학작품을 문단, 문장, 단어 등으로 분리하여 계층적으로 선으로 연결된 꽃잎의 형태로 표현한 종래의 텍스트 시각화 결과를 나타내고,  
 도 5는 종래의 한국의 족보를 시각화한 그림을 나타내고,  
 도 6은 본 발명의 실시예에 따른 시 텍스트의 시각화 방법의 흐름을 나타낸 도면이고,  
 도 7은 본 발명의 실시예에 따른 시 텍스트 시각화 방법에 적용되는 L 시스템을 이용한 가지의 생성 예를 나타내는 모식도이고,  
 도 8은 본 발명의 실시예에 따른 시 텍스트 시각화 방법에 적용되는 구조적 매핑의 모식도이고,  
 도 9는 본 발명의 실시예에 따른 시 텍스트 시각화 방법에 적용되는 내용의 내용적 매핑으로서, 단어의 특성과 유사한 색을 나타내는 모식도이고,  
 도 10은 본 발명의 실시예에 따른 시 텍스트 시각화 방법에서 시 중복도와 나무의 모양을 나타낸 모식도이고,  
 도 11은 본 발명의 실시예에 따른 시 텍스트 시각화 방법을 적용해 이상의 '거울'을 시각화한 결과를 나타내고,  
 도 12는 이상의 '오감도'의 시제1호를 시각화한 결과를 나타내고, 도 13의 이상의 '오감도'의 시제2호를 시각화한 결과를 나타내고,  
 도 14는 본 발명의 실시예에 따른 시 텍스트 시각화 방법을 적용하여 김소월의 '진달래꽃'을 시각화한 결과를 나타내고, 도 15는 김소월의 '반달'을 시각화한 결과를 나타내며, 도 16은 김소월의 '산유화'를 시각화한 결과를 나타낸다.

### 발명을 실시하기 위한 구체적인 내용

[0027] 본 발명의 이점 및 특징, 그리고 그것을 달성하는 방법은 첨부되는 도면과 함께 상세하게 후술되어 있는 실시예들을 통해 설명될 것이다. 그러나 본 발명은 여기에서 설명되는 실시예들에 한정되지 않고 다른 형태로 구체화될 수도 있다. 단지, 본 실시예들은 본 발명이 속하는 기술분야에서 통상의 지식을 가진 자에게 본 발명의 기술적 사상을 용이하게 실시할 수 있을 정도로 상세히 설명하기 위하여 제공되는 것이다.

[0028] 도면들에 있어서, 본 발명의 실시예들은 도시된 특정 형태로 제한되는 것이 아니며 명확성을 기하기 위하여 과장된 것이다. 또한 명세서 전체에 걸쳐서 동일한 참조번호로 표시된 부분들은 동일한 구성요소를 나타낸다.

[0029] 본 명세서에서 "및/또는"이란 표현은 전후에 나열된 구성요소들 중 적어도 하나를 포함하는 의미로 사용된다. 또한, 단수형은 문구에서 특별히 언급하지 않는 한 복수형도 포함한다. 또한, 명세서에서 사용되는 "포함한다" 또는 "포함하는"으로 언급된 구성요소, 단계, 동작 및 소자는 하나 이상의 다른 구성요소, 단계, 동작, 소자 및 장치의 존재 또는 추가를 의미한다.

[0030] 이하에서 본 발명의 바람직한 실시예를 도면을 참조하여 상세히 설명하기로 한다.

[0031] 도 6은 본 발명의 실시예에 따른 시 텍스트의 시각화 방법의 흐름을 나타낸 도면이다. 도 6에 나타난 바와

같이, 본 발명의 실시예에 따른 시 텍스트 시각화 방법은, (a) 시각화 장치에 입력된 시의 텍스트를 순서대로 스캐닝하는 단계(S100); (b) L 시스템 모델을 통해 상기 시의 구조를 식물의 구조와 매칭시키고, 상기 텍스트의 형태소 및 모음을 식물의 채도 및 명도와 매칭시켜 컴파일링 하는 단계(S200); (c) 컴파일된 상기 텍스트의 명령어에 따라 렌더링하여 페인팅하고, 상기 식물의 그림을 생성하는 단계(S300); 및 (d) 생성된 상기 그림을 상기 시각화 장치의 표시창에 표시하는 단계(S400)를 포함한다.

[0032] 이처럼 본 발명은 컴퓨터 알고리즘을 사용하여 시의 패턴을 분석하고 분석된 패턴을 알고리즘 기반 모델링을 이용하여 기능적이면서 동시에 예술적 이미지로 시각화하는 한 방법을 제안하고 실험적 연구 결과를 개시한다. 즉, 본 발명은 시의 특징을 잘 드러내는 한 그루의 나무로 표현하고, 이때 나무는 L-System을 사용하여 컴퓨터 알고리즘에 의해 자동으로 생성하는 것을 제안한다. 이렇게 시의 언어적 코드를 디지털 공간으로 매핑 하여 창조된 그림은 시를 다른 공간에서 변형하여 바라봄으로써 감상하는 데 도움을 주면서도 동시에 또 다른 예술작품의 원천으로 사용되는 것을 보는 즐거움을 준다.

[0033] 또한 발명에서는 한국의 대표 시인 김소월과 이상의 시를 대상으로 한 시각화 실험 결과는 제안하는 방법이 정보 전달 면에서 기능적이고 동시에 심미적인 결과를 만들어 낼 수 있음을 보여주었다. 텍스트를 비평적으로 읽는 기존의 방법과 더불어 이러한 텍스트의 예술적 시각화 변형 방법은 그 엔터테인먼트적 요소 때문에 더 많은 독자의 관심을 불러 일으켜 한국 시에 대한 더 깊은 담론이 이루어지게 할 수 있으며 기존의 친숙한 한국 시를 재조망하게 해 주며 또한 한국어를 모르는 외국인이나 시를 어려워하는 이들이 쉽게 한국 시에 다가갈 수 있게 할 수 있는 가능성을 열어 준다는 면에서 큰 장점이 있다.

[0034] 그리고, 본 발명의 실시예에 따른 시 텍스트 시각화 방법을 수행하기 위한 시각화 장치는 PC, 태블릿 PC, 휴대폰, PDA 등 프로그램을 수행하여 결과를 표시하는 장치이면 모두 가능하다.

[0035] 도 6에 나타난 바와 같이, 먼저 시각화 장치에 텍스트로서 시가 입력되면, 시각화 장치는 텍스트를 스캐닝하고 (S100), 스캐닝된 텍스트를 컴파일링하고(S200), 컴파일된 텍스트 명령어를 렌더링시켜 페인팅하여 그림을 생성한 후(S300), 시각화 장치의 표시창에 상기 그림을 표시하게 된다.(S400)

[0036] 여기서, 본 발명의 실시예에 따른 시 텍스트의 시각화 방법에서 (b) 단계는(S200), L 시스템 모델을 통해 시의 구조를 식물의 구조와 매칭시키고, 상기 텍스트의 형태소 및 모음을 식물의 채도 및 명도와 매칭시켜 컴파일링하는 단계이다.

[0037] 먼저, 본 발명의 실시예에 따른 시 텍스트 시각화 방법에 적용되는 L 시스템을 살펴보면, L-System은 1968년에 Aristid Lindenmayer 라는 사람에 의해 개발된 문자열 재작성 시스템으로 1984년에 Alvy Ray Smith에 의해 컴퓨터 그래픽에 도입되었다. 그 후 주로 식물의 성장 등을 절차적으로(procedural) 모델링하는 데 유용하게 쓰이고 있다.

[0038] L-System은 심볼들과 심볼들을 대체하는 규칙으로 이루어져 있다. 예를 들어,

[0039] 
$$F \rightarrow F [+F] F$$

[0040] 위와 같은 규칙은 현재의 문자열에서 모든 'F' 심볼을 " $F [ + F ] F$ " 로 대체하는 규칙이다. 최초 문자열이 'F' 였다면 위와 같은 규칙을 두 번 적용하면 다음과 같은 문자열로 확장되게 된다.

[0041] 
$$F \rightarrow F [+F] F \rightarrow F [+F] F [+F [+F] F] F [+F] F$$

[0042] 이러한 L-System에서 각 심볼에 기하학적인 의미를 부여하면 이러한 문자열은 알고리즘적으로 그림을 그리는 일련의 명령들로서 작용하게 된다. 예를 들어 위의 문자열상의 각 심볼들에 다음과 같은 기하학적 의미를 부여한다면,

[0043] F : 현재 위치에서 오른쪽으로 ' ' 그리기

[0044] + : 45도 시계 반대방향으로 방향을 바꾸기

[0045] [ : 현재 위치 저장하기

[0046] ] : 마지막에 저장된 위치로 돌아가기

[0047] 위의 문자열은 두 번 반복하여 규칙이 적용되는 과정을 통해 다음과 같은 모양을 만들어 내게 된다. 이렇게 L-System은 생물의 자기 복제 시스템을 모방할 수 있는 방법으로 식물의 알고리즘기반 모델링에 매우 유용하게 사용되어 지고 있다. 도 7은 본 발명의 실시예에 따른 시 텍스트 시각화 방법에 적용되는 L 시스템을 이용한 가지

의 생성 예를 나타내는 모식도이다. 그러므로, 본 발명은 이와 같은 식물의 알고리즘 기반 모델링(L 시스템)을 이용하여 텍스트로 이루어진 시를 나무로 시각화하는 방법을 제안하고자 한다.

[0048]

보다 구체적으로, 생물학적 설명에 따르면 인간의 시각인지 시스템은 두 가지로 나뉜다. 첫째는 'where' 시스템으로 사물의 움직임이나 위치, 깊이, 전체적인 구조를 인지하는데 사용되어지는데 이 시스템에서 색 정보는 사용되어지지 않고 밝기 정보만이 사용되어 진다. 둘째는 'what' 시스템으로 사물을 인식하는데 사용되어 지는데 이때는 색 정보가 중요하게 이용되어 진다. 이러한 생물학적 특성을 근거로 제안하는 방법에서는 전체적으로 보았을 때 시의 구조를 나무의 형태적 구조로 매핑시키고 내용을 잎사귀의 색으로 시각화하는 방법을 취한다.

[0049]

#### 시 구조의 시각화

[0050]

본 발명에서는 시의 시각적 재해석을 위해 나무라는 메타포를 사용하는 이유는 우선적으로 시의 계층 구조를 매우 잘 나타내 줄 수 있기 때문이다. 또한, 시인은 시를 창작할 때 연과 행과 어절, 그리고 글자의 패턴으로 운율을 만든다. 이러한 시의 시각적 계층 구조는 작가가 나타내고자 하는 바를 담는 그릇이라고 할 수 있는데 몸통과 가지와 잎사귀로 이루어지는 나무는 본질적으로 그러한 계층적 구조를 매우 잘 나타낼 수 있는 가능성을 가지고 있다.

[0051]

도 8은 본 발명의 실시예에 따른 시 텍스트 시각화 방법에 적용되는 구조적 매핑의 모식도이다. 도 8은 정지용 시인의 ‘호수1’이라는 시를 예제로 시의 계층적 구조를 구성하는 시, 연, 행, 단어가 어떻게 나무의 형태로 매핑이 되는지 보여준다.

[0052]

도 8에 나타난 바와 같이, 나무의 몸통은 시 자체를 표현한다. 한 연은 몸통에서 나오는 가지로 표현이 된다. 이때 일련의 연들은 시계 방향으로 회전하여 나오는 가지들로 표현이 된다. 위 예제에서는 연이 두개이기 때문에 몸통에서 나오는 두 개의 가지로 표현이 된다. 각 연에서 각 행은 더 가는 가지들로 표현이 된다. 연과 마찬가지로 일련의 행들은 시계방향으로 회전하며 나오는 가지들로 표현이 된다. 따라서 위 예에서 첫 번째 연은 삼행으로 구성되어 있기 때문에 세 개의 가는 가지로 표현된 것을 볼 수 있다.

[0053]

각 행은 또한 다수의 어절들로 구성이 되어 있어 더 가는 가지들과 잎사귀를 이용하여 마찬가지로 방법으로 시간 방향 순서로 돌아가며 표현이 된다. 위 예제에서 첫째 연의 첫째 행은 두 어절로 구성이 되어 있으므로 잎사귀가 붙은 두 개의 가장 가는 가지가 뻗어 나온 것을 볼 수 있다.

[0054]

마지막으로 각각의 어절은 다수의 글자로 이루어져 있고 각 어절의 크기, 즉 한 어절을 이루는 글자 수는 잎사귀의 두께로 매핑이 된다. 즉 적은 글자 수로 이루어진 어절은 가는 잎사귀로 반대의 경우 굵은 잎사귀로 표현이 된다. 위 예제의 마지막 행에서 한 글자로 이루어진 ‘수’ 어절을 나타내는 잎사귀가 두 글자로 이루어진 ‘밖에’에 해당하는 잎사귀보다 가는 잎사귀로 표현된 것을 볼 수 있다. 이처럼 시를 구성하는 연, 행, 어절, 글자의 계층 구조가 시각 도메인에서 색상을 제외하고 형태로만 매핑이 되어 표현되어 지게 된다.

[0055]

#### 시 내용의 시각화

[0056]

나무의 형태가 시의 구조를 나타낸다면 사계절에 따라서 색을 달리하는 잎사귀는 시의 내용을 표현하기에 적절하다. 도 9는 내용의 내용적 매핑으로서, 단어의 특성과 잎사귀 색을 나타내는 모식도이고, 정지용님의 ‘호수1’ 시에서 두 번째 연에 대해서만 구조뿐 아니라 내용까지 시각화한 결과를 보여준다.

[0057]

이처럼 본 발명의 실시예에서는 시의 내용을 구성하는 많은 요소 중 컴퓨터로 분석이 비교적 용이한 형태소의 품사와 각 음절이 가지는 밝음과 어두움을 시의 내용을 구성하는 요소로 사용하였다.

[0058]

먼저 형태소의 품사의 경우, 명사나 조사 같은 품사는 딱딱한 느낌을 만들어 내므로 낮은 채도(gray)의 색을 갖는 잎사귀로 표현이 되며 동사나 형용사 등은 그 반대의 느낌을 나타내는 높은 채도의 색을 갖는 잎사귀로 표현이 되게 된다. [표 1]은 형태소의 품사가 채도로 어떻게 매핑이 되는지 보여준다. 이때 한 어절이 여러 형태소로(여러 품사)로 이루어져 있는 경우 그 평균값을 사용하게 된다.

[0059]

예를 들면 도 9에 나타난 바와 같이, ‘마음’이나 ‘호수’는 체언으로 채도가 없는 회색으로 표현이 되어 있는 반면 ‘보고’나 ‘싶은’ 같은 어절은 품사가 용언이므로 채도가 높은 녹색으로 표현되어 진 것을 볼 수 있다.

표 1

| 형태소 품사               | 채도 (0-1) |
|----------------------|----------|
| 체언(명사, 수사, 대명사), 조사  | 0        |
| 용언(동사, 형용사), 부사, 감탄사 | 1        |
| 그 외                  | 0.5      |

[0060]

[0061]

한편, 한글의 모음은 양(밝음)과 음(어두움), 그리고 중성적인 성격을 가지는 모음들로 나눌 수 있다. 이러한 발음상의 특징은 시를 읽으며 감상하는 자에게 일정한 느낌의 차이를 만들어 낸다. 이러한 성질을 이용하여 [표 2]와 같이 각 모음을 잎사귀 색의 밝기로 시각적으로 매핑한다. 한 어절에 해당하는 한 잎사귀의 밝기는 그 어절을 구성하는 모음들의 명도 평균값을 사용하여 결정되어 진다.

표 2

| 모음                              | 명도 (0-1) |
|---------------------------------|----------|
| ㅡ, ㅣ, ㅏ, ㅑ, ㅓ, ㅕ, ㅗ, ㅛ, ㅜ, ㅠ, ㅡ | 0        |
| ㅜ, ㅛ, ㅕ, ㅗ, ㅛ, ㅜ, ㅠ, ㅛ          | 1        |
| 그 외                             | 0.5      |

[0062]

[0063]

도 9에 나타난 바와 같이, 어절 ‘보고’ 나 ‘하니’ 는 밝은 모음들로 이루어져 밝은 잎사귀로 표현되는 반면 ‘눈’ 이나 ‘싶은’ 같은 어절은 모음이 어두움 느낌을 가지기에 어두운 잎사귀로 표시되는 것을 보여 준다.

[0064]

명도와 채도는 어절의 특성에 따라 바뀌지만 잎사귀 간 색 차이를 보다 명확하게 드러내기 위해 모든 잎사귀의 색(hue값)은 하나의 녹색값으로 동일하게 설정하여 표현되었다.

[0065]

한편, 본 발명의 실시예에서는 각 어절의 특징을 개별 잎사귀로 표현하는 것과 더불어 전체적인 내용의 중복도를 나무 몸통과 가지의 굵기로 시각적으로 표현한다. 시는 정보 전달 목적의 텍스트와 달리 표현의 수단으로 중복을 사용한다. 이러한 시의 특성을 나타내기 위해 각 시당 중복도를 계산한다. 중복도 계산을 위해서 각 어절 j당 중복도 S를 [수학식 1]과 같이 정의한다.

### 수학식 1

$$S_j = \frac{1}{n} \sum_{i \neq j} dist(어절_i, 어절_j), 0 \leq dist(.) \leq 1$$

n = 한 시에 포함된 어절의 개수 - 1  
 $dist(어절_i, 어절_j)$  = 서로 다른 어절간 유사도

[0066]

[0067] 따라서, 시 당 중복도 S는 다음과 같이 계산한다.

## 수학식 2

$$S = \frac{1}{n} \sum S_j$$

(n = 한 시에 포함된 어절의 개수)

[0068]

[0069] 어절과 어절 사이의 유사도 dist(.)는 문자열 간의 유사도 계산에 많이 사용되는 Levenshtein distance를 이용하여 계산하게 된다. 이렇게 계산된 중복도는 나무의 몸통과 가지의 굵기로 매핑되어 도 10에 나타난 바와 같이, 비슷한 어절이 중복해서 사용된 시는 굵은 나무로 시각화된다. 도 10은 본 발명의 실시예에 따른 시 텍스트 시각화 방법에서 시 중복도와 나무의 모양을 나타낸 모식도이다.

## [0070] 구현 알고리즘

[0071] 상술한 본 발명의 실시예에 따른 시 텍스트 시각화 방법은 L-System을 이용하여 알고리즘 적으로 생성된다. 시 시각화를 위한 전체 알고리즘 프로세스는 도 6에 나타난 프로세스와 동일하다.

[0072] 즉, 입력된 시는 컴파일이라는 과정을 통해 그림을 그리기 위한 일련의 명령어로 바뀌게 되며 페인트 단계에서 실제 그림으로 렌더링 되게 된다. 컴파일 과정에서는 시의 어절을 순서대로 스캐닝하면서 나무를 생성시킬수 있는 일련의 명령어들을 회귀적(recursive)으로 만들어 내게 된다. 이것을 위해서 본 발명에서는 L-System을 사용함에 있어 다음의 규칙을 정의하였다.

[0073]  $F \rightarrow F [ + F ] \text{ or } F [ -F ]$

[0074] 즉, F 가 시 전체라고 한다면 한번 L-System을 적용할 때 한 연에 해당하는 가지가 왼쪽 또는 오른쪽 비스듬한 방향으로 번갈아 가며 뻗어나가게 된다. 한편, 시 내용을 명시하기 위해서 각 잎사귀의 채도와 명도, 그리고 크기는 0-9 레벨로 표현되어 L-System 문자열에 인코딩된다. 가령 채도 6, 명도 3, 그리고 크기 4 레벨의 잎사귀는 ‘;’ 심볼을 사용하여 다음과 같이 표현된다. 즉, ‘;’ 심볼이 붙은 ‘F’ 는 모두 잎사귀로 표현이 되며 그렇지 않은 ‘F’ 는 가지로 표현이 된다.

[0075] [ ;634F ]

[0076] 이와 같은 방법으로 컴파일된 시는 마지막 렌더링 단계에서 각 심볼이 기하학적으로 해석되며 Bezier Curve를 사용하여 자연스럽게 나무 모양으로 시각화 된다.

## [0077] 이상과 김소월 시의 알고리즘 기반 시각화

[0078] 상술한 본 발명의 실시예에 따른 알고리즘을 이용한 시 텍스트 시각화 방법에 따라 한국을 대표하는 두 시인의 대표작들을 대상으로 제시한 방법을 적용시켜보았다. 이상과 김소월은 비슷한 시대에 활동한 한국을 대표하는 시인이지만 서로 매우 다른 성향을 가진 시인으로서 그들의 대표작들을 시각화함으로써 제시한 방법이 각 시인의 성향과 작품의 특색을 적절히 시각화하여 새로운 감상과 재해석의 틀로서 기능한지 확인하려고 하였다. 실험에서 형태소 분석을 위해서 Komoran 자바 한글 형태소 분석기(v. 1.12)를 사용하였고 시각화를 위하여 Processing(<http://processing.org/>)을 이용하였다.

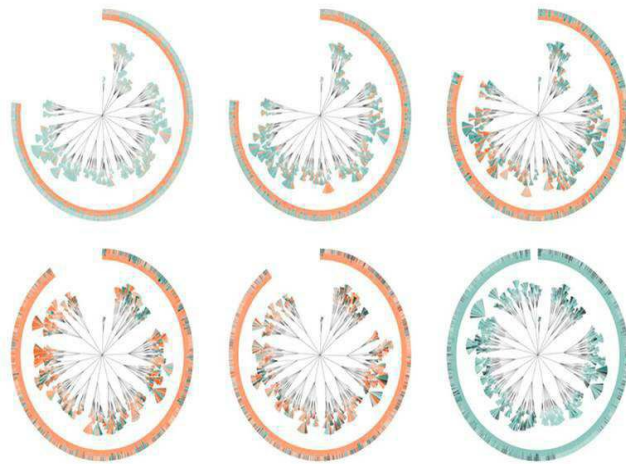
[0079] 알고리즘을 이용한 시각화에 앞서 작품의 인쇄 원본 자체가 일종의 시각화라는 사실을 인지하는 것이 중요하다. 시가 최초 선보일 때 독자는 시각적으로 작품을 읽고 감상하기에 시각적 배치를 신경 써서 만들기 때문이다. 따라서 본 발명의 실시예에 따른 시각화를 위해 작품이 최초 실린 일제시대 원본을 구입하여 사용하였다.

- [0080] 하지만 본 실험을 위해 사용하고 있는 텍스트 분석기가 일제시대에 사용되었던 언어 분석이 어려워 시각적 배치(예, 띄어쓰기)가 아닌 용어 자체는 현재 사용되어지는 용어로 바꾸어 사용하였다. 이를 위해 김소월시전집(권영민 엮음, 문학 사상사, 2007)과 이상전집 1(권영민 엮음, 서울:뿔, 2009)을 사용하였다.
- [0081] 이상(1910-1937)은 일제 강점기에 활동한 한국의 대표적인 시인이다. 많은 분야의 연구자들이 이상을 연구하였는데-수학자가 본 이상(김명환 교수), 시각디자인 연구자가 본 이상(김민수 교수), 정신분석학자가 본 이상(주두영 교수), 철학자가 본 이상(김상환 교수)-그 만큼 이상의 시는 다양한 해석과 감상의 여지를 준다. 그 당시 조선미술전람회에도 출품할 정도로 그림에도 능했던 이상의 시는 재미 비평가 Walter Lew에 따르면 반복과 확산, 정체성과 억압, 욕구와 소멸이라는 대조적인 패턴을 시각적으로 언어적으로 자신의 시를 통해 능숙하게 탐색한다.
- [0082] 도 11은 본 발명의 실시예에 따른 시 텍스트 시각화 방법을 적용해 이상의 '거울'을 시각화한 결과를 나타내고, 도 12는 이상의 '오감도'의 시제1호를 시각화한 결과를 나타내고, 도 13의 이상의 '오감도'의 시제2호를 시각화한 결과를 나타낸다.
- [0083] 도 11 내지 13에 나타난 바와 같이, 시각화된 이상의 시들은 마치 외계에서 온 나무처럼 생소하다. 본 시각화 알고리즘은 원본 인쇄본에서 어절 간 존재하는 시각적 간격을 기본으로 나무를 생성시킨다. 따라서 띄어쓰기를 표현의 수단으로 사용한 이상의 시들은 매우 특이한 형태의 나무로 생성이 된다.
- [0084] 도 11은 이상의 대표작인 '거울'을 제안한 방법에 의해 시각화한 결과를 보여 주고 있다. 이 시의 하나의 특징은 띄어쓰기에서 찾아 볼 수 있는데, 도 11에서 그러한 특징이 나무 잎사귀의 두께로 잘 표시되고 있음을 볼 수 있다. 즉, 한 단어를 이루는 글자 개수가 많음으로 인해 잎사귀들의 두께가 매우 두껍게 표현되어 짐을 볼 수 있다.
- [0085] 도 12는 이상의 '오감도' 시제1호를 제안한 방법에 의해 시각화 결과를 보여 주고 있다. 시제1호는 비슷한 구절을 반복적으로 사용하고 있는 것이 가장 큰 특징인데 따라서 중복도 계수가 크게 계산되고 그러한 특징이 그림에서 보이듯이 나무 기둥과 가지의 굵기로 표현되고 있다. 즉, 나무 기둥과 가지가 다른 시 보다 굵게 표현되어 지고 있는데 이것은 중복에 의해 강조되는 표현을 시각적으로 감상하게 도와준다.
- [0086] 도 13은 이상의 '오감도' 시제2호를 시각화한 결과를 보여주는데, 이상 시가 지니는 극단적인 형태적 특이함을 시각적으로 잘 보여주고 있음을 볼 수 있다. 즉, 연과 행과 단어(띄어쓰기에 의해 구분되는)가 모두 하나밖에 없는 이 시는 하나밖에 없는 앙상한 가지에 하나의 넓은 잎사귀로 표현된다. 이렇듯 이상의 시의 형태적 특징과 내용적 특징(반복)이 시각화를 통해 직관적으로 동시에 심미적으로 표현되고 있음을 볼 수 있다.
- [0087] 모던니즘과 전통은 동전의 양면과 같다. 이상의 시가 모던니즘을 대표한다면 김소월(1902-1934)의 시는 보통 전통적으로 여겨진다. 이상의 시가 독자로 하여금 시를 읽은 후 끊임없는 의문을 가지게 만들어 마치 이국적인 땅에 있는 서 있는 것 같은 느낌을 가지게 만든다면 김소월의 시는 그와 매우 다른 느낌을 선사한다.
- [0088] 도 14는 본 발명의 실시예에 따른 시 텍스트 시각화 방법을 적용하여 김소월의 '진달래꽃'을 시각화한 결과를 나타내고, 도 15는 김소월의 '반달'을 시각화한 결과를 나타내며, 도 16은 김소월의 '산유화'를 시각화한 결과를 나타낸다.
- [0089] 도 14 내지 도 16에 나타난 바와 같이, 김소월의 시는 앞서 이상의 시와 매우 대조적인 결과를 보여 준다. 매우 이국적인 이상의 나무와 달리 김소월의 나무는 자연스럽고 친숙하다. 이상의 시들과 비교하였을 때 김소월의 나무는 대체적으로 무성한 가느다란 가지에 가느다란 잎사귀들로 이루어져 있음을 본다. 이것은 시 자체에 여백이 많고 중복해서 쓰이는 어구도 적다는 것을 의미한다. 또한 이러한 형태적인 차이뿐만 아니라 잎사귀의 색도 이상의 것들과 비교한다면 훨씬 그 밝기와 채도의 변화가 다양하다. 그것은 김소월의 시들이 좀 더 다양한 타입의 언어를 사용하였음을 간접적으로 암시한다. 또한, 김소월 나무들은 이상과 달리 대체적으로 비슷한 모습을 보이는 데 이것은 시 창작에 있어 모던니스트적인 실험을 했던 이상과 구별되어지는 작가의 특징을 간접적으로 보여 준다고 할 수 있겠다.

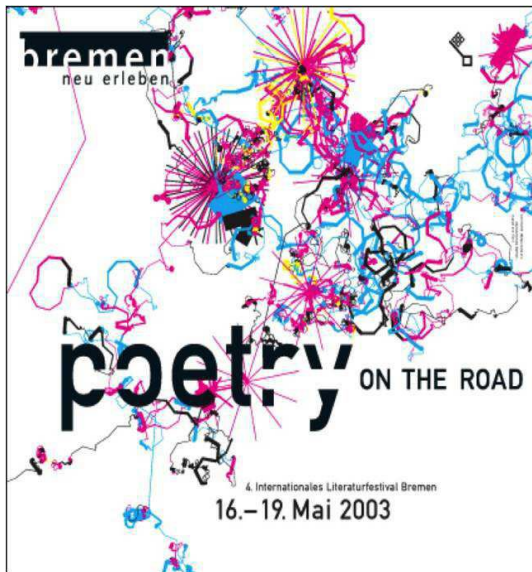
- [0090] 이와 같이, 본 발명에서는 시를 변형(deformance)하여 새로운 방향에서 심미적으로 감상하게 도와주는 방법을 제안하고 그 실험결과를 개시한다. 제안하는 방법은 한국 시에 내재되어 있는 패턴을 컴퓨터 등의 시각화 장치를 이용해 발견해 내고 그러한 언어적 패턴을 컴퓨터 알고리즘을 이용해 디지털 캔버스 상에서 일련의 심미적인 형태와 색으로 재현해 낸다.
- [0091] 즉, 한국 시의 언어적 코드를 디지털 공간으로의 매핑을 통해 그림으로 창조해 낸다. 제안하는 알고리즘에 의해 생성되어지는 메타포적인 이미지들은 정보를 제공함과 동시에 심미적이라는 큰 장점이 있다.
- [0092] 그리고 시각화를 통해 이러한 심미적/엔터테인먼트적 요소는 몇 가지 의의를 가진다. 첫째, 기존의 알고 있던 시를 전혀 다른 도메인에서 재창조시켜 보는 것은 색다른 감상의 즐거움을 주며 원문을 새로운 각도에서 감상하게 만든다는 점과, 둘째, 어려운 시 이어서 접근하기 힘들었거나 한국시를 잘 모르는 외국인에게는 그림을 통해 쉽게 원문 시로 관심을 유도할 수 있는 가능성도 보여 준다는 점과, 셋째, 이러한 심미적 시각화 결과는 거꾸로 새로운 시의 창작으로의 영감을 부여하는 가능성을 보여 준다는 점이다.
- [0093] 이상의 설명에서 본 발명은 특정의 실시 예와 관련하여 도시 및 설명하였지만, 특허청구범위에 의해 나타난 발명의 사상 및 영역으로부터 벗어나지 않는 한도 내에서 다양한 개조 및 변화가 가능 하다는 것을 당 업계에서 통상의 지식을 가진 자라면 누구나 쉽게 알 수 있을 것이다.

## 도면

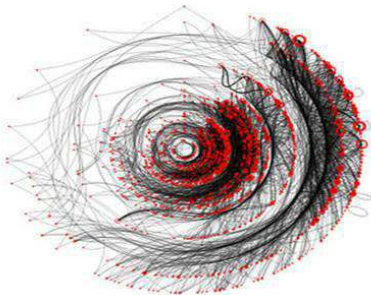
### 도면1



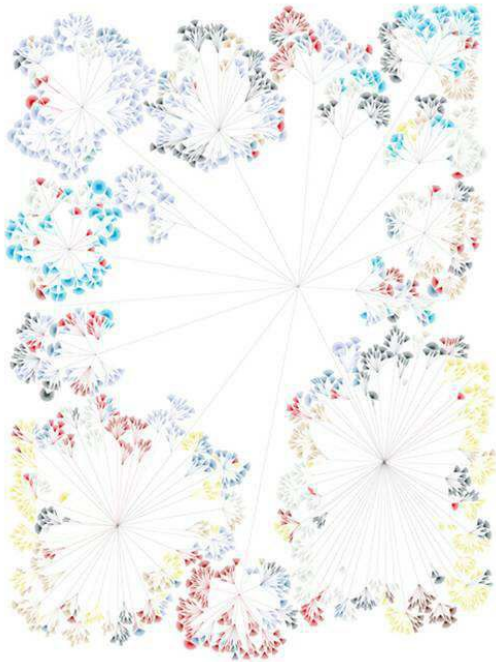
도면2



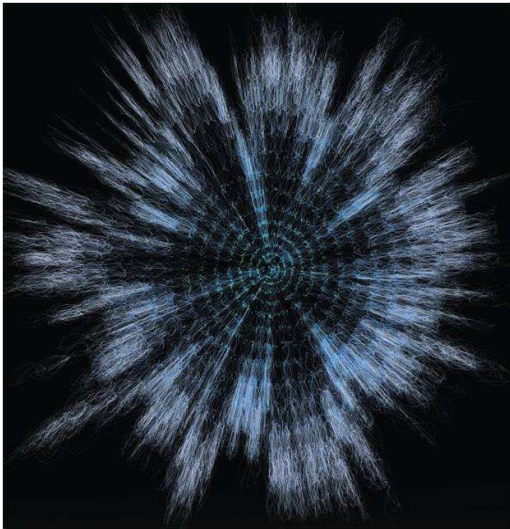
도면3



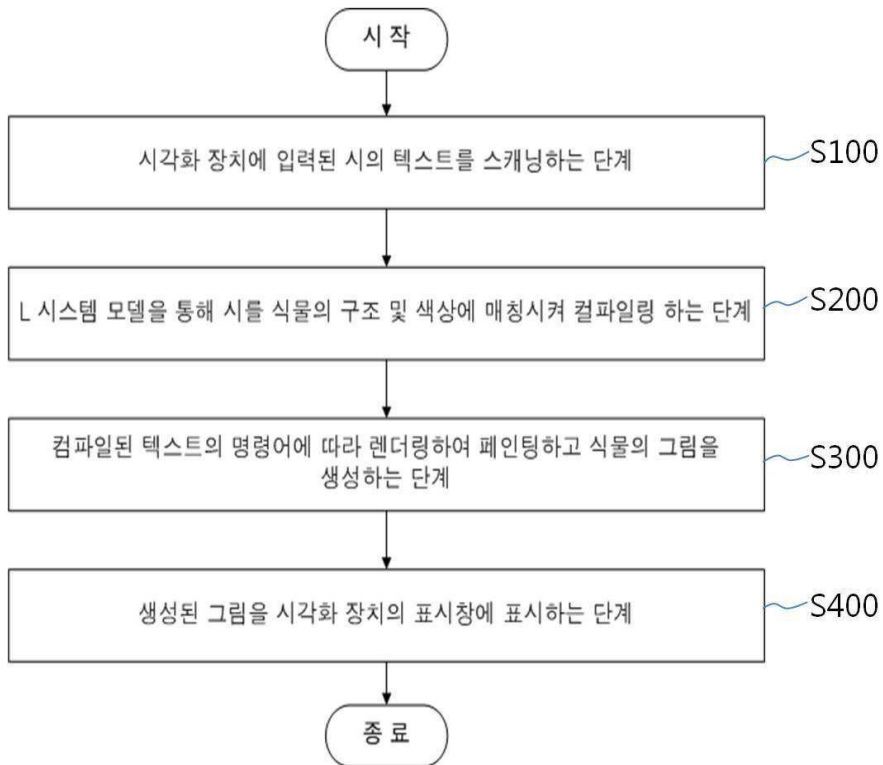
도면4



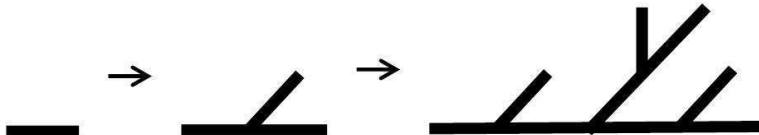
도면5



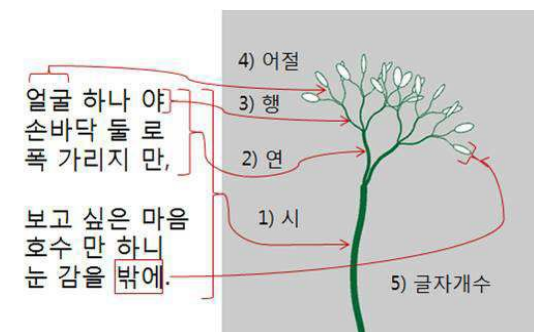
도면6



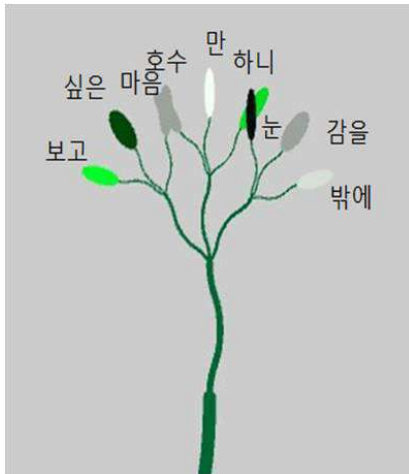
도면7



도면8

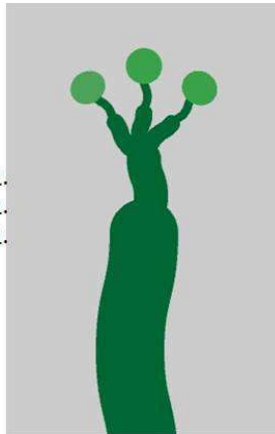


도면9



도면10

제1의아해가무섭다고그리오.  
제2의아해도무섭다고그리오.  
제3의아해도무섭다고그리오.



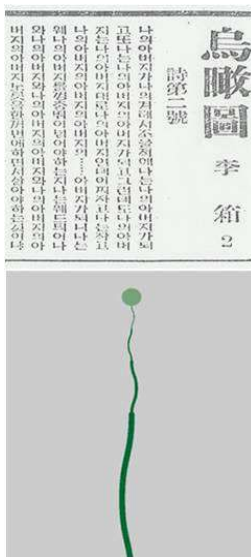
도면11



도면12



도면13



도면14



도면15



도면16

