

K 엔진 백서

목차

서문4

1. K 엔진의 근본원리.....	6
(1) 정보의 저장단위.....	6
(2) 체계.....	7
(3) 연결정보.....	8
2. K 엔진의 구체적 작동방식.....	9
(1) 교과서.....	9
(2) 지식단위.....	10
(3) 위치정보.....	11
i. 부모목차 정보.....	11
ii. 레벨정보.....	12
iii. 동일레벨에서의 자신의 순서정보.....	12
(4) 연결정보.....	12
3. 태그기준의 다양성.....	16
4. 인간의 사유구조와 K 엔진.....	17
5. K 엔진의 확장.....	18
(1) 수직적 확장.....	18
(2) 수평적 확장.....	18
(3) 연결로직의 확장.....	19
6. GPT 와의 비교.....	20
7. 관련특허.....	20
Introduction.....	24

1. Fundamental Principles of K-Engine.....	26
(1) Storage Unit of Information.....	26
(2) System.....	27
(3) Linked information	27
2. How K-Engine Works	29
(1) Textbook.....	29
A textbook looks roughly like this.....	29
(2) Knowledge Unit.....	29
(3) Location Information	31
i. Table of Contents of Parent	31
ii. Level information	32
iii. Own order information on the same level.....	32
(4) Linked information	32
3. Diversity of Tag Criteria	36
4. Human Structure of Thinking and K-Engine	37
5. Expansion of K-Engine	37
(1) Vertical Expansion.....	38
(2) Horizontal Expansion	38
(3) Expanding Connecting Logic	39
6. Comparison with GPT	40
7. Related Patents	41

서문

K엔진은 완전히 새로운 인공지능이다. 단지 대화형 인공지능의 범주를 넘어 인류지식 전체를 총망라한다.

개인적으로 인공지능의 개발을 시작한 지 어언 20년이 흘렀다. 필자의 지난 20년은 오로지 이 프로그램 하나만을 개발하는 데 바쳐졌다고 해도 과언이 아니다. 코어 부분에 관한 현재 1.0버전이 완성된 이 시점에서 뒤를 돌아보면 이리도 간단한 것이 왜 그리도 막막하게만 보였는 지 새삼 신기하기 그지 이를 데 없다.

모든 것이 만들어지고 나면 단순하다. 만들어진 후에야 단순하다.

지난 20년간 하등의 보수도 받는 것이 없이 오로지 개인 시간과 돈을 투입하여 탄생시킨 것이 K엔진이다. K엔진은 Knowledge 엔진이란 뜻이지만 Korea Engine 라고 생각해도 좋다. 이름은 아무래도 좋다.

언제나 항상 했던 생각이 하나 있는데 그것은 K엔진과 같은 존재는 대기업에서 수백억의 연봉을 받는 엔지니어 머릿속에서는 절대 나올 수 없다는 점이다. 인간은 자신이 받는 만큼만 노동한다. 수백억을 받으면 기껏해야 수백억짜리만 생산해 낼 뿐이다. 수백조를 주는 기업이 없으므로 수백조짜리는 그 어떤 기업의 개발실에서도 잉태될 수 없다. 그런 물건은 언제나 창고나 차고 같은 곳에서만 탄생하게 된다.

현재 글쓰는 시점에서 K엔진의 포장지 버전은 0.6이다. 특히 자연어검색 부분이 60% 정도의 정확도 밖에 없다. 데이터 역시 법인세의 60% 정도(700페이지 중 400페이지) 정도 밖에 올라와 있지 않다. 그러나 코어 버전은 당당히 말하건대 1.0 이다. 보름 후 기자회견 때에는 데이터는 100% 그 외 나머지 포장지 부분이 80% 정도까지는 향상이 이루어질 것으로 예상된다.

개인적으로 하고 싶은 것만 하고 산 사람으로서 지금까지 주변 사람들에게 참으로 많은 피로감을 안겨준 것만큼은 사실이다. 내가 길바닥에 나섰던 700명의 뮤리츠 주주들, 내 부모, 나의 형제와 나의 처자식들 나에게 돈을 투자한 몇분의 투자자들 모두에게 감사드린다.

2023.09.25.
능평동 사무실에서

1. K엔진의 근본원리

(1) 정보의 저장단위

컴퓨터에서 데이터는 하드디스크에 저장된다. 그리고 메모리가 이를 끌어와서 CPU에 집어넣어 하고 싶은 일을 하고 다시 밖으로 내뱉는다. 이것이 튜링의 컴퓨터이다. 여기에는 저장장치가 있다. 하드디스크에는 정보(데이터)가 저장된다. 주로 언어로 저장된다. 그 언어를 이리저리 가공한 것이 챗GPT나 구글 바드와 같은 대화형 인공지능이다.

여기서 두가지 눈여겨 볼 장면이 있다. 하나는 정보가 “언어” 데이터로 저장된다는 사실이다. 다른 하나는 데이터들간의 상호연결에 관한 이렇다 할 처리기준이 없다는 사실이다. 컴퓨터가 바보인 이유는 바로 이 때문이다.

현재의 컴퓨터는 정보를 “언어”로 저장한다. 그리고 그 언어들을 이리저리 가공하여 2차 가공물을 생산해 낸다. 여기서 언어는

“주민등록증”, “번호”, “이름”, “주소”

등과 같은 것을 의미한다. 현재의 데이터의 저장 및 처리 단위는 이와 같은 언어, 특히 단어이다.

이렇게 저장된 정보는 서로간의 연결관계에 관한 정보가 없다. 없을 수 밖에 없다. 도대체 망고와 국가 사이에 어떤 연결관계가 있을 수 있으며 사랑과 여행 사이에 어떤 뚜렷한 연결관계를 미리 규정할 수 있겠는가? 그래서 데이터들은 서로 뚝 떨어져서 따로들 놀게 된다. 그리고 이렇게 서로 무관한 데이터들을 어떻게든지 조합하여 뭔가 내뱉어 보겠다는 시도가 바로 GPT다.

그러니 엄청나게 고가인 CPU 내지 GPU를 주문해야 하고 이와 동시에 엄청난 전기 또한 소모해야 한다. 그리고 그 결과에는 거짓말도 꽤나 섞여서 나오게 된다.

근본적으로 컴퓨터는 스스로 무슨 말을 하는 지 자체를 모르고 있다. 이는 마치 제어봉이 빠져 버린 핵원자로와 같다. 굉음을 내면서 돌아갈 뿐 언제 터져버릴 지 모르는 시한폭탄과도 같다. 이것이 GPT이며 이 하자는 DNA 차원에서 존재하는 치명적인 불치병이다. 이것이 GPT가 영원한 양치기 소년에서 벗어날 수 없는 이유다.

(2) 체계

K엔진에서는 정보가 “지식”의 단위에서 저장된다. 예를 들어

“주민등록증에는 이름, 주민등록번호, 주소가 기재되어 있다.”

와 같다. 이것이 하나의 지식단위이다. 그리고 이러한 지식들은 “체계”라는 상호연결규칙으로 서로 연결되어 있다. 또한 이들 지식들이 체계적으로 한데 뭉친 것을 두고 일반적으로 “교과서”라고 부른다. 교과서는 보통 책이라는 오프라인 물체로 존재한다. 반면 K엔진에서는 이 세상의 모든 지식들을 “온라인 교과서”화 시킨다. 즉 K엔진은 이 세상의 모든 지식에 대한 온라인 교과서이다.

보통 교과서는 A4 1,000페이지 정도이다. 평균적으로 그렇다는 뜻이다. 이 정도라면 예를 들어 민사소송법이라는 어느 한 과목에 대한 고시 답안지를 생성해 내는 데는 충분하다. 예전에는 사법시험이나 행정고시와 같은 시험이 있었는데 이러한 시험이라면 무조건 전국 수석합격을 할 수 있는 정도가 되는 것이 1,000페이지 정도의 분량이다. 다만 실무적으로 사용하려면 약 5,000 ~ 10,000페이지 정도는 되어야 하고 이 정도면 변호사와 바둑으로 친다면 호선으로 대적하게 된다. 나아가 100,000페이지 정도까지 나아가게 되면 변호사가 두점이나 석점 정도를 깔고 두어야 형세가 만만하게 된다.

교과서가 교과서인 이유는 해당 분야에 관한 모든 답변을 할 수 있기 때문이다. 이것이 교과서의 개념이다. 교과서가 교과서 대접을 받기 위해서는 해당 분야의 모든 지식들을 총체적으로 망라하여 쓸어 담아야 한다. 그러자니 모든 중복을 제거하고 질서정연하게 지식들을 정렬시켜야 한다. 그 정렬방법이 바로 “체계”이다. K엔진은 체계엔진이다. 이 체계를 바탕으로 모든 질문에 대한 답변을 한다. 사법시험을 수석합격한 원희룡이 즉시 답변해 준다. K엔진이 데이터를 대략 100,000페이지 정도 소화시키고 나면 원희룡이 두점이나 석점 정도 깔고 두는 수준까지 이르게 된다. 이러한 시대는 대략 2025년도 말쯤이면 도래하게 된다. 즉 이 세상의 모든 전문직이 사라지는 날은 앞으로 2년 정도 밖에 남지 않았다. 그날은 어느 순간 도둑처럼 다가온다.¹

¹ 내가 도둑 같이 이르리니(요한계시록 3장 3절). 그 외 데살로니가전서 5장 2절, 마태복음 24장 30절 등등. 참고로 필자는 크리스찬은 아니다. 굳이 답변한다면 힌두교 쪽에 가깝다.

(3) 연결정보

언어는 독립적으로 존재하지만 지식은 체계적으로 존재한다. 바로 이 차이가 GPT와 K엔진의 차이이다. GPT는 데이터(언어) 간의 연결정보가 없지만 K엔진은 그것을 가지고 있다. K엔진의 정보들은 언제나 타 정보들과의 연결정보와 함께 데이터베이스에 저장된다. 또한 동시에 스스로의 체계상의 위치정보를 보유하고 있다. 컴퓨터는 이 두가지 정보, 즉 연결관계정보와 체계정보를 바탕으로 전체 지식데이터들간의 상호연관관계를 이미 파악하고 있다. 그리고 질문이 들어오면 원희룡보다 빠르고 정확하고 친절하게 답변해 준다. 더욱 중요한 것이 한가지 있는데 그것은 돈을 받지 않는다는 점이다.

현재까지 데이터들 간의 연결정보는 사실상 존재하지 않았다고 볼 수 있으며 이는 K엔진에 의하여 이 세상에 최초로 도입된 것이라 할 수 있다. 일반적인 데이터베이스에 있어서 DBMS라는 것이 있기는 하지만 이것은 지식간 연결관계가 아니라 테이블간의 연결관계, 즉 일종의 행렬식의 연결관계일 뿐이다. 하나의 지식 단위에서 다른 모든 지식과의 연결관계라는 개념은 컴퓨터나 데이터베이스의 역사상 한 번도 존재하지 않았던 개념이다. 특히 GPT에게 이 개념이 결여되어 있다. GPT는 단어와 단어 사이의 확률적 연결관계만을 추론할 뿐 의미적 연결관계를 추론하지 못한다. 반면 K엔진은 근본적으로 “의미적 연결관계”에서 출발하고 있다.

“체계”와 “연결정보”. 이것이 바로 K엔진이다.

2. K엔진의 구체적 작동방식

(1) 교과서

교과서는 대략 다음과 같이 생겼다.

목 차		ix
제 2 장 민사소송의 기본개념과 기본원리		
제1절	민사소송의 구조	40
I.	재판의 구조	40
II.	심리의 구조	41
1.	청구(본안의 신청)의 단계	42
2.	법률상 주장 단계	43
3.	사실상 주장 단계	43
4.	증명(입증)의 단계	45
제2절	소송의 주체	45
제1항	법 원	45
I.	법원의 의의	45
II.	법원의 구성	45
1.	합의제	46
2.	단독제	46
3.	합의제와 단독제의 조화	46
III.	법관의 제척·기피·회피	47
1.	의 의	47
2.	제 척	47
3.	기 피	49
4.	회 피	54
제2항	당사자	54
I.	당사자의 의의	54
1.	형식적 당사자개념	54
2.	당사자의 호칭	55
3.	당사자는 민사소송의 주역-당사자권	55
4.	두 당사자 대립의 구조	55
II.	당사자의 확정	56
1.	의 의	56
2.	당사자확정의 기준	56
3.	당사자확정이 문제되는 영역	57

교과서를 대표하는 것이 그 책의 목차이다. 그리고 목차는 그 교과서가 지닌 체계를 대변한다. 교과서는 이런 식으로 이 세상의 모든 지식들을 한 권의 책 속에 질서있게 정렬시킨다. 그리고 질문이 있을 때마다 교과서의 전체 체계 속에서 필요한 부분을 추출하고 다시 이를 체계화시켜 답변을 제공한다. 따라서 잘 만들어진 교과서라면 적어도 자신의 분야에 관한 한 “모든” 답변이 가능하다. 무한답변이 가능하다. 이것이 교과서이고 이것이 체계이

며 이것이 바로 K엔진이다.

(2) 지식단위

교과서에 담겨 있는 것은 지식이다. 절대적인 것은 아니지만 지식은 보통 1개의 문장에 담겨있다. 즉 1지식에 1개의 문장이 할당된다. 그리고 그 1개의 지식에는 한 개의 제목이 붙어있다. 즉 제목(subject)과 내용(content)의 쌍으로 이루어진 것이 하나의 지식단위가 된다.

K엔진은 교과서의 모든 지식들을 이 두가지 방식으로 재정리한다. 왜냐하면 그래야만 컴퓨터가 그 지식을 먹고 소화시킬 수 있기 때문이다. 일단 컴퓨터가 지식을 섭취하고 소화시키고 나면 컴퓨터는 이제 인간이 된다. 변호사가 되고 의사가 되고 ... 된다. 이 세상의 모든 분야의 전문가가 된다. 앞서도 한 번 언급했지만 각 분야마다 10,000페이지 정도의 지식만 섭취하고 나면 이제 이 세상의 전문가는 모두 사라지게 된다. 극히 보조적인 역할만을 하는 정도로만 남게 된다. 10,000페이지는 100명이 100일이면 입력이 가능한 분량이다. 10명이면 1,000일, 즉 3년이다. 결국 2년 정도면 이 세상의 모든 지식들이 컴퓨터 속으로 빨려 들어가게 된다. 지상천국은 그리 멀지 않은 곳에 있다.

지식단위는 지식 그 자체를 나타내는 문장과 그 지식이 다른 지식들과 관련하여 가지고 있는 각종 정보들로 구성되어 있다. 즉 제목, 내용, 연결정보가 지식단위를 구성하는 일단의 기본요소가 된다. 교과서 그 자체가 가지고 있는 체계정보에 상응하는 위치정보까지 여기에 더하여짐으로써 지식단위의 4대 구성요소가 완성된다. 그 하나의 예를 들면 다음과 같다. 일단 그 이전에 다음과 같은 상상을 미리 해 보자. 즉 클라이언트가,

“법인으로 보는 단체로 승인받은 경우 거주자 변경 가능 여부”

라는 키워드 내지 문장으로 K엔진에게 질문을 던졌다고 하자. 그러면 통상 다음의 17번 지식이 검색된다.

1. 제목
 - (1) 거주자 또는 비거주자로 변경 제한
2. 내용
 - (1) 법인으로 보는 단체로 승인을 받은 경우 승인일이 속하는 과세기간과 그 과세기간이 끝난 날로부터 3년이 되는 날이 속하는 사업연도까지는 거주자 또는 비거주자로 변경할 수 없습니다. 다만, 세무서장 승인 요건을 갖추지 못하게 되어 승인 취소된 경우는 예외입니다
3. 위치정보
 - (1) 전체체계상의 위치정보
 - i. 법인세신고 > A. 법인세의 납세의무 > iii. 법인구분별 납세의무 > (3) 비영리 법인의 납세의무 > 2) 비영리법인의 종류 > 다. 국세기본법 §13의 법인으로 보는 단체 > B. 세무서장의 승인에 의해 법인으로 보는 단체
 - (2) 위치정보 아이디
 - i. 17
4. 연결정보
 - (1) e3<n>,r13<이상에서와 같이>

이상에서 제목과 내용에 대한 설명은 필요없으며 위치정보와 연결정보, 즉 태그(tag)에 대해서만 설명하기로 한다.

(3) 위치정보

위치정보는 해당 지식이 전체 체계 속에 위치하는 좌표정보라 할 수 있다. 두가지 형태로 표현될 수 있는데 그 중 하나는 전체 체계와의 관계를 모두 나타내는 방식이며 다른 하나는 그 위치를 아이디화 한 위치정보이다.

위치정보는 다시 3가지로 구분된다.

i. 부모목차 정보

체계적 지식은 수형도 형태로 이루어져 있으며 수형도는 부모자식관계를 바탕으로 이루어져 있다. 따라서 수형도 상에 존재하는 그 어떤 지식단위도 자신의 부모가 누구인지만

알면 자신의 체계상의 위치를 확인할 수 있게 된다.

ii. 레벨정보

이는 실은 부모정보를 파악하는 과정에서 자연스럽게 확정되는 정보이다. 모든 수형도 상의 지식들은 상위단위와 하위단위로 체계화되며 이 경우 하위단위는 직상위단위에 대하여 한단계 레벨이 낮아지게 된다. 1. 아래에 1) 이 오는 경우 1) 은 1. 보다 한 레벨 아래에 위치하게 된다. 출력물 상에서는 오른쪽으로 탭(Tap)을 한번 치는 모습으로 나타나게 될 것이다.

iii. 동일레벨에서의 자신의 순서정보

이는 자신이 위치한 레벨에서 동위의 다른 지식단위들이 동등하게 위치하고 있는 경우 그들간의 관계에서의 순서정보이다. 즉 형제관계에서의 서열번호이다. 즉 동일레벨에서의 1, 2, 3, ... 또는 A, B, C, ... 와 같은 목차를 통하여 나열되는 경우 자신이 B에 해당한다면 그 B가 자신의 순서정보가 된다.

(4) 연결정보

연결정보는 해당 지식이 다른 지식들과의 연결관계를 나타내는 정보이다. 위 사례에서는 e3와 r13의 두개의 태그가 존재한다. (참고로 ◇속의 글자들은 13번 지식과 연결이 되는 경우 접속어를 의미한다. ‘n’은 접속어가 필요없음을 의미한다.)

e3은 3번 지식의 필수(Essencial)적 동반을 의미한다. 3번 지식은 다음과 같다.

1. 제목
 - A. 세무서장의 승인에 의해 법인으로 보는 단체
2. 내용
 - A. 일정한 유형의 법인으로서는 수익을 구성원에게 분배하지 아니하는 것은 법인으로 보아 세법을 적용합니다.
3. 위치정보
 - A. 전체 체계상의 위치정보
 - i. 생략
 - B. 위치정보 아이디
 - i. 15

그리고 r13은 권장(Recommended) 동반지식을 의미한다. 13번 지식은 다음과 같다.

1. 제목
 - (1) 요건
2. 내용
 - (1) 사단, 재단, 그 밖의 단체의 조직과 운영에 관한 규정을 가지고 대표자 또는 관리인을 선임하고 있을 것.
 - (2) 사단, 재단, 그 밖의 단체 자신의 계산과 명의로 수익과 재산을 독립적으로 소유·관리할 것.
 - (3) 사단·재단 그 밖의 단체의 수익을 구성원에게 분배하지 아니할 것.
3. 위치정보
 - (1) 전체 체계상의 위치정보
 - i. 생략
 - (2) 위치정보 아이디
 - i. 13

보통 1개의 지식은 1개의 문장으로 나누어지지만 경우에 따라서는 위와 같이 3개의 매우 밀접한 관련을 가진 지식을 하나로 통합하여 하나의 지식으로 삼는 경우도 있다. 그 결과

이상의 세가지 지식을 통합한 결과는 다음과 같다.

1. 세무서장의 승인에 의해 법인으로 보는 단체
일정한 유형의 법인으로서 수익을 구성원에게 분배하지 아니하는 것은 법인으로 보아 세법을 적용합니다.
2. 요건
 - (1) 사단, 재단, 그 밖의 단체의 조직과 운영에 관한 규정을 가지고 대표자 또는 관리인을 선임하고 있을 것.
 - (2) 사단, 재단, 그 밖의 단체 자신의 계산과 명의로 수익과 재산을 독립적으로 소유·관리할 것.
 - (3) 사단·재단 그 밖의 단체의 수익을 구성원에게 분배하지 아니할 것.
3. 거주자 또는 비거주자로 변경 제한
이상에서와 같이 법인으로 보는 단체로 승인을 받은 경우 승인일이 속하는 과세기간과 그 과세기간이 끝난 날로부터 3년이 되는 날이 속하는 사업연도까지는 거주자 또는 비거주자로 변경할 수 없습니다. 다만, 세무서장 승인 요건을 갖추지 못하게 되어 승인 취소된 경우는 예외입니다

이 최종의 결과가 최초의 제17번 지식이 검색되었을 경우 K엔진이 출력해 내는 결과이다.² 하지만 여기서 끝나지만은 않는다. 위 결과는,

“법인으로 보는 단체로 승인받은 경우 거주자 변경 가능 여부”

라는 질문에 대해 하나의 지식, 즉 17번 지식만이 추출되는 모습만을 설명하고 있지만 실제로는 수십개의 지식들이 추출되는 것이 보통이다. 즉 위 사례는 그 중 가장 정확도가 높았던 17번 지식만을 바탕으로 한 출력결과에 지나지 않는다.

K엔진은 통상 상위 3 ~ 4 개의 검색결과를 바탕으로 위와 같은 과정을 거친 출력을 발생시킨다. 즉, rank 점수가 100점인 피검색지식이 17번이고 85점인 피검색지식이 34번이고 또한 77점이 피검색지식이 5번이라 하고 그들의 연결정보가 다음과 같다고 한다면,

² 위 결과에서 '이상에서와 같이'라는 접속사가 추가되었음이 나타나고 있다.

- 7번 지식
 - rank : 100
 - 연결정보(tag) : e3<n>,r13<이상에서와 같이>
- 34번 지식
 - rank : 85
 - 연결정보(tag) : e22<n>,r3<n>, r19<n>
- 5번 지식
 - rank : 77
 - 연결정보(tag) : r27<n>,r35<n>, r41<n>

최종적으로 추출되는 지식데이터는 다음과 같다.

e3, r3, **5**, r13, **17**, r19, e22, r27, 34, r35, r41, **77**

굵은 숫자는 원천적으로 추출된 지식데이터들이고 그 외는 이들을 바탕으로 연결정보를 통하여 동반적으로 추출된 지식데이터들이다. 이 중 e3와 r3은 중복되므로 한번만 출력되면 된다. 그리하여 최종적으로 출력되는 아이디의 결과집합은,

3, 5, 13, 17, 19, 22, 27, 34, 35, 41

의 총 10개의 지식이 된다. 만약 필수동반지식들만을 추출하여 노출시킨다면

3, 5, 17, 22, 77

의 총 5개의 지식들만이 노출되게 될 것이다.

이들 지식들은 모두 고유의 체계상의 위치정보를 보유하고 있다. 그 위치정보는 다음 세가지로 구분된다. 즉 앞서 언급한,

- 부모목차정보
- 레벨정보
- 동일레벨에서의 자신의 순서정보

가 그것이다. K엔진은 이 정보를 바탕으로 추출된 지식들에게 다시 새로운 목차를 부여하면서 최종 결과를 질서정연한 모습으로 출력해 낸다.

3. 태그기준의 다양성

앞서 연결정보를 나타내는 태그기호들(e3<n>,r13<이상에서와 같이> - 7번지식에 대한 연결정보)은 해당 피검색 지식데이터(즉 7번 지식)를 기준으로 하여 그 지식과 연결관계에 있는 지식들을 수작업으로 추출한 것이다. 이 때 그 추출의 기준은 “의미 자족성”이었다. 즉 해당 피검색 지식데이터(7번) 하나만 추출하게 되면 이 지식이 어떤 문맥에서 등장하고 있는 지 불분명할 뿐만 아니라 심지어는 아예 해당지식의 의미 자체마저 파악하기 힘들 수 있다. 이에 이러한 부족함을 보완해 줄 수 있는 보조 지식데이터들이 필요하게 된다. 즉 이 경우의 동반지식데이터의 추출기준은 “의미의 자체완결성”이라 할 수 있다.

문제는 “의미의 자체완결성”은 동반지식데이터 추출의 유일한 기준이 아니라 무수한 기준 중 하나라는 사실이다. 앞서와 같이 단지 의미의 자체완결성이 아닌 또 다른 기준, 예를 들면 “실제 사례를 바탕으로 해당 실제사례와 관련된 동반지식”을 추출할 수도 있다. 이렇게 되면 일단 무수한 실제사례를 먼저 수집한 후 그 하나하나마다 이와 연관된 동반지식을 추출하는 방식으로 태그를 구성할 수도 있다. 이는 일종의 “질문의 다양한 관점”에 대한 “답변의 다양한 관점”이라 할 수 있다. 그리고 그 관점은 다차원적인 형태로 결합될 수 있다. 즉 앞서의 가장 기본적인

이 때 “의미의 자체완결성”을 그 기초로 하게 된다면³

1. 의미의 자체완결성
2. 실제사례 관련성

이라는 두가지 차원으로 동반지식을 추출할 수도 있다.

좀 더 쉽게 설명한다면 음식에 대한 질문에서 영양소 관점에서의 질문이 있고 다이어트 관점, 가격 관점, 그 외 다양한 관점에서의 질문이 있을 수 있다. 이에 대한 답변 역시 그 동반지식의 결과집합이 달라질 수 밖에 없다. 그리고 이러한 관점 중 가장 기본이 되는 것이 의미의 자체완결성이라 할 수 있다.

이처럼 단지 일차원적인 추출 뿐만 아니라 다차원적인 추출까지 하게 된다면 이는 인간의 두뇌수준을 현격하게 뛰어넘게 된다. 인간의 두뇌 용량이나 스피드는 다만 1차원적인 동반지식 추출을 하는 정도로도 벅차기 때문이다. 그러나 컴퓨터는 2차원, 3차원, 4차원, ...

³ 대부분의 경우 실제로 그렇게 될 수 밖에 없다. 의미의 자체완결성은 가장 기본적인 연결정보 추출 기준이라 할 수 있기 때문이다.

용량이 받쳐주는 범위 내에서 무한차원까지 도달할 수 있는데 이 정도면 가히 신의 등장이라 하지 않을 수 없게 된다.

4. 인간의 사유구조와 K엔진

“체계”는 이 세상의 모든 지식을 담을 수 있다. 지금까지는 이것으로서 충분했다. 지금까지는 책이라는 유형물에 모든 지식을 담을 수 밖에 없었고 그리하여 그 지식들을 체계화하는데 까지만 작업이 가능했고 더 이상의 어떤 처치를 할 방법도 없었다. 이것이 지난 수 천년 동안 “책”이라는 형태로 지식이 정비되고 전수되고 이해되고 교육되어왔던 방식이다.

어떤 문제되는 상황이 발생하면 일단은 책을 뒤진다. 물론 그 책이 머릿속에 잘 소화되어 정리되어 있는 사람, 즉 변호사와 같은 사람은 책을 뒤질 필요 없이 자신의 머릿속의 두뇌 책을 사용하면 된다. 이 두뇌책의 속도는 엄청 빠르다. 즉 컴퓨터와 마찬가지로이다. 질문을 받은 변호사는 즉각 답변한다. 이는 두뇌속에 저장된 정보들이 뇌신경망을 통해 서로 고도의 속도로 통신한 결과이다.

이 장면에서 변호사의 머릿속 어딘가에는 지식들간의 연결정보가 저장되어 있음을 추측할 수 있다. 이 연결정보들이 상호 교통하면서 멋드러진 질서정연한 답변이 입에서 술술 흘러나오는 것이다. 이것을 구현한 것이 K엔진이다.

이상에서 알 수 있듯이 지금까지는 오프라인 상에 “책”이라는 유형물을 만들어 내는 것에 그치고 그 책 속의 지식들간의 연결관계는 - 각고의 노력을 투입하여 - 인간의 머릿속 뇌 신경세포에 저장했다고 할 수 있다. 즉 그 두가지는 분리되어 저장되었다.

이들 분리된 두가지 요소를 통합한 것이 바로 K엔진이다. 컴퓨터는 이 구가지 요소를 하나의 하드디스크에 저장한다. 즉 이제 책만 만들면 더 이상 각고의 노력을 할 필요가 없어진 것이다. 책만 만들어지면 이제는 공부를 하지 않고도 누구나 변호사가 되는 것이다. 의사도 그렇고 약사도 마찬가지이다. 이 세상의 모든 전문지식이 “공부”할 필요 없이 나의 것이 되는 것이다. 나아가 구글링을 통해 얻는 일반정보 역시 체계화가 된다. 그리하여 일상적인 지식을 얻는 영역 역시 K엔진의 지배권 하에 들어오게 된다.

일단 인간의 수작업에 의한 연결정보가 입력되기만 하면 그 다음에는 교과서의 체계적 특성에 바탕하여 모든 질문에 답변을 할 수 있다. 이 때 그 수작업은 내가 하는 것이 아니라 남이 해 준다. 나는 아무것도 할 것이 없다. 공부 안해도 언제나 100점이다. 학교 갈 이유가 없어진다.

5. K엔진의 확장

현재 이 책이 쓰여지는 시점에서 단지 700페이지짜리 세법, 그 중에서도 법인세 분야 하나만의 입력작업이 이루어지고 있다. 여기서 두가지 방향에서의 확장이 이루어지게 된다.

(1) 수직적 확장

개인적으로 필자는 앞서도 언급한 것처럼 궁극적으로 분야당 100,000페이지 정도의 데이터 입력을 예상하고 있다. 이렇게 되면 아무리 공부를 수십, 수백 년 한 변호사보다도 실력이 좋아지게 된다. 그 어떤 세세한 내용이라 할지라도, 그 어떤 복잡하고 깊이있는 내용이라 할 지라도 K엔진은 이를 족집게처럼 집어낸다. 변호사나 그 외 전문가들은 다시 자료를 뒤지는 과정을 거쳐야 하지만 K엔진은 그런 과정 없이 즉각 답변을 제공해 준다. 단 한번의 입력만으로 인류 모두가 변호사가 되고 의사가 된다. 다만 업데이트는 필요하다. 그리하여 기존 변호사나 의사들은 업데이트 입력으로 그 역할을 이어가게 된다.

K엔진의 원리가 단순한 것처럼 보여도 그 단순함은 이 세상의 모든 지식을 담을 수 있는 단순함이다. 단순할수록 정확하다. 단순할수록 에러 위험이 없다. 단순할수록 무한하다. 단순할수록 보다 고도로 진화할 수 있다. K엔진의 진화는 무한하다.

필자의 생각으로는 이제 더 이상의 K엔진은 등장하기 힘들 것으로 보인다. 설령 등장한다면 지라도 그 필요가 없다. 지금의 윈도우가 존재하는 이상 새로운 OS를 특별히 개발할 필요가 없는 것과 같다. 즉 윈도우가 PC OS의 종착역이듯이 K엔진 역시 K엔진의 종착역이지 않을까 싶다. K엔진은 더 이상 단순화시킬 수 없을 정도로 단순하기 때문이다. 더 이상의 특허가 나오는 것도 불가능해 보인다.

결론적으로 K엔진은 대화형 인공지능의 결정판이며 더 이상은 없다. 앞으로의 대화형 인공지능의 성장은 K엔진 외부가 아니라 K엔진 내부에서 이루어지게 된다.

(2) 수평적 확장

필자는 K엔진을 지주회사로 하여 일단 1,000개의 법인을 설립하고 이들 모두를 나스닥이나 코스닥에 상장시킬 계획이다. 개별적으로 라이선스를 부여하여 라이선스 갱신을 바탕으로 하는 경쟁을 유도할 계획이다. 라이선스를 받은 기업이 게으름을 노정한다면 재계약을 해 주지 않고 다른 경쟁 희망기업에 라이선스를 부여하게 된다.

K엔진의 기초데이터의 입력은 1분야당 10억원 정도로 여겨진다. 1,000개 기업이라면 1조

원 정도가 소요되게 된다. 필자는 이를 직접 실시할 생각은 없지만 적어도 라이선스를 바탕으로 지배권은 행사할 생각이다.

시간이 흐를수록 분야는 확장되고 세분화될 것이다. 대략 100,000개 정도의 분야로 세분화될 것으로 예상된다. 내가 태어난 이 나라 한국 정부의 입장에서 본다면 한 분야당 평균적으로 10명씩만 고용창출이 이루어진다면 총 1,000,000개의 일자리가 창출될 것이며 이때 대한민국의 1인당 GDP는 세계 최고가 될 것으로 예상된다. 국가 GDP 마저도 최고가 될 수 있지 않을까 하는 추측도 해 본다.

(3) 연결로직의 확장

K엔진의 발전이 앞으로 무궁무진한 이유는 바로 이 곳에 있다. 이는 앞서 언급한 태그기준, 즉 연결로직의 다양화와 그 고차원화에 기반한다. 이로써 K엔진은 수평적으로 모든 분야로 확장되고, 그 개개의 분야마다 수직적으로 고도로 깊이를 더해가며 마지막으로 다양한 관점의 변경(variation)을 통하여 인류의 지식체계를 초고도화 시키게 된다. 이렇게 되면 인조인간이 진정한 인간이 되고 지금의 인간이 인조인간 수준으로 전락할 수 밖에 없게 된다. 나는 나의 자녀가 나보다 더 훌륭한 존재가 되기를 바란다. 마찬가지로 K엔진이 인간의 손으로 만들어졌지만 인간보다 더 훌륭한 존재가 되기를 바란다.

마지막으로 질문 하나를 던진다. 지금까지의 설명 정도면 K엔진이 인공지능의 결정판, 그 최종적인 프로토콜이 될 수 밖에 없다는 사실을 거의 증명한 것으로 생각된다. 이의있는가?

6. GPT와의 비교

비교 항목	GPT	K엔진
작동기반	언어	지식
기초데이터	웹상의 정보	교과서적 지식
데이터베이스 용량	빅데이터 기반	스몰데이터 기반 ⁴
원천자료로부터의 오류 상 속 가능성	있음	없음 ⁵
지식의 조합과정에서의 오 류 발생 가능성	있음	없음
전문지식에 대한 답변	취약	강력
일반지식에 대한 답변	강력	단기적 취약, 장기적 강력 ⁶
답변의 깊이	얕음	깊음
저작권 문제	발생	없음 ⁷
서버사양	고사양 필요	고사양을 필요로 하지 않음
하드디스크 사양	초대형 용량 필요	한 분야당 100MB정도 충분
전력소비량	거대	일반 서버(구글서버 등)와 동일 ⁸

7. 관련특허

- 대한민국 특허출원 10-2023-0053098 : 지식교차와 지식연결을 통한 연결지식 생성 방법 및 시스템(Connection Knowledge Generating Method and System Through Knowledge Crossing and Knowledge Connection)
- 대한민국 특허출원 10-2023-0108659 : 동반지식의 추출을 통한 결합지식의 생

⁴ 어느 한 분야에서 변호사급 답변을 하는데 필요한 데이터량은 10메가바이트 정도이다.(약 10,000페이지)

⁵ 물론 교과서의 제작 자체에서 오류가 발생한다면 K엔진 역시 오류가 발생할 수 있다. 물론 이런 경우는 이론적으로만 가능한 이야기일 뿐이다.

⁶ 일반지식 역시 장기적으로는 모두가 K엔진에 담기게 된다. 이 시기가 오면 일반지식 분야에서도 K엔진이 GPT를 능가하게 된다.

⁷ GPT는 인터넷 상의 지식들을 모두 끌어모으는 과정에서 필연적으로 저작권 이슈가 발생하지만 K엔진은 스스로 데이터(교과서)를 제작하므로 저작권 문제로부터 완전 자유롭다.

⁸ 구글보다 한층 작은 전력을 사용하게 된다. 그 이유는 구글 역시 빅데이터 기반이기 때문이다. K엔진은 에센스화 된 스몰 데이터를 기반으로 로직으로만 승부하므로 소요 전력량은 극히 작다.

성방법(Method of generating combined knowledge through extraction of accompanying knowledge)

WHITEPAPER OF K-ENGINE

Introduction

K-Engine is a completely new artificial intelligence. It is not just a category of interactive artificial intelligence, but goes beyond the human knowledge.

It has been 20 years since I personally started developing AI. It is no exaggeration to say that the last 20 years of my life have been devoted to developing this program alone. Regarding the core, we are now at version 1.0, and looking back, it's amazing how something so simple seemed so daunting at the time.

Everything is simple once it's built. It's only simple after it's made.

For the past 20 years, I have been working on the K-engine during my personal time and spending money without any compensation. K-engine means Knowledge Engine, but you can also think of it as Korea Engine as well. I like the name.

I've always thought that something like K-Engine could never come from the minds of engineers earning tens of billions of dollars in large corporations. People only work for what they are paid. If you're paid tens of billions, you're only going to produce tens of billions at best. Trillions can't be conceived in the development of any company because no company gives out trillions. This kind of work will always start in a warehouse or garage.

At the time of writing this, the wrapper version of K-Engine is 0.6. In particular, the natural language search is only about 60% accurate. The data is also only about 60% of corporate taxes (400 pages out of 700 pages). But I'm proud to announce that the core version is 1.0. By the time of the press conference in two months, the data will be 100% and the rest of the wrapper will be improved up to 80%.

As someone who has lived a life of doing what I'm eager to do, I can tell you that I have caused a great deal of exhaustion to those around me. Thank you to the 700 Muritz shareholders who had to bear the loss, my parents, my brothers, my wife and children, and the few investors who invested money into my project.

September 25, 2023
From my office in Neungpyeong-dong

1. Fundamental Principles of K-Engine

(1) Storage Unit of Information

In a computer, data is stored on the hard disk, and the memory pulls it in, puts it into the CPU, does what it wants to do, and puts it out again. This is Turing's computer. It has storage. The hard disk stores information (data). It is mainly stored as a language, and the processing of the language is the interactive AI such as ChatGPT or Google Bard.

There are two things to note here. One is the fact that information is stored as "language" data. The other is the fact that there are no processing standards for the interconnection of data. This is why computers are fools.

Current computers store information as "languages." They process those languages to produce secondary products. In this case, language means:

"Social security card", "Number", "Name", "Address", etc.

The current unit of data storage and processing is language, specifically words.

The information stored in this way has no information about how they relate to each other. It is impossible. How could there be a connection between a mango and a country, or how can we definitely make a connection between love and travel in advance? Thus, the data is left to play in isolation. And GPT is an attempt to somehow combine these unrelated data and come up with something.

Therefore, you have to order an incredibly expensive CPU or GPU, and at the same time, you have to consume a lot of electricity. And the result includes a bunch of lies as well.

Basically, the computer doesn't know what it's talking about. It's like a nuclear reactor with a control rod missing. It's a ticking time bomb that's spinning and spinning, but it doesn't know when it's going to explode. This is GPT, and this flaw is a fatal incurable disease that exists at the DNA level. This is why GPT will always be a lying shepherd boy.

(2) System

In the K-engine, information is stored in units of "knowledge". For example:

"A national identity card contains a name, social security number, and address."

This is a unit of knowledge. And this knowledge is connected to each other by an interconnection rule called "system." This systematic collection of knowledge is commonly referred to as a "textbook." Textbooks usually exist as offline objects called books. In contrast, K-Engine makes all the knowledge in the world into an "online textbook." Therefore, K-Engine is an online textbook of all the knowledge in the world.

A typical textbook is about 1,000 A4 pages. That is, on average. That's enough to generate the answer key for a single subject, for example, civil procedure law. In the past, there were exams such as the judicial exam and the administrative exam in Korea, and you need 1,000 pages, which is enough to pass the national exam with the top grade. However, for practical use, it should be about 5,000 to 10,000 pages, and if we think of this as an example of a playing "Go" with a lawyer, you will have the first move in the alternate game. Furthermore, if you know 100,000 pages, the lawyer will have to lay down two or three points to play a fair game

The reason why a textbook is a textbook is because it can give you all the answers in the field. That's the idea of a textbook. In order for a textbook to be considered a textbook, it has to be a comprehensive compilation of all the knowledge in the field. To do so, it must eliminate all redundancies and arrange the knowledge in an orderly fashion. The organizing method it is called "system." K-Engine is a system engine. Based on this system, all questions are answered. Won Hee-ryong, who passed the bar exam with a top grade, answers immediately. After K-Engine has digested about 100,000 pages of data, it will reach a level where Won Hee-ryong can put down two or three points. This era is expected to arrive by the end of 2025, which means that the day when all professional jobs in the world will disappear is only about two years away. That day will come like a thief.⁹

(3) Linked information

Language exists independently, but knowledge exists systematically. This is the difference between GPT and K-Engine. GPT has no linked information between data

⁹ I will come like a thief (Revelation 3:3). 1 Thessalonians 5:2, Matthew 24:30, etc. For your information, I am not a Christian, and I am more of a Hindu.

(language), but K-Engine has it. Information in K-Engine is always stored in a database with links to other information, and it also has its own systematic location information. The computer already knows the interrelatedness of all the knowledge data based on these two pieces of information, namely the linked information and the system information. And when asked a question, it answers faster, more accurately, and more kindly than Won Hee-ryong. More importantly, it doesn't get paid.

Until now, the linked information between data has been virtually non-existent, and it can be said that it was first introduced to the world by K-Engine. In a typical database, there is a DBMS, but it is a connection between tables, which is merely a kind of matrix linkage, not a connection between knowledge. The idea of a connection from one unit of knowledge to all other units of knowledge has never existed in the history of computers or databases. In particular, GPT lacks this concept. GPT can only infer probabilistic connections between words and words, not semantic connections. On the other hand, K-Engine is fundamentally based on "semantic connections".

"System" and "linked information". This is K-Engine.

2. How K-Engine Works

(1) Textbook

A textbook looks roughly like this.

제 2 장 민사소송의 기본개념과 기본원리	
제1절 민사소송의 구조	40
Ⅰ. 재판의 구조	40
Ⅱ. 심리의 구조	41
1. 청구(본안의 심청)의 단계	42
2. 법률상 주장 단계	43
3. 사실상 주장 단계	43
4. 증명(입증)의 단계	45
제2절 소송의 주제	45
제1항 법 원	45
Ⅰ. 법원의 의의	45
Ⅱ. 법원의 구성	45
1. 합의제	46
2. 단독제	46
3. 합의제와 단독제의 조화	46
Ⅲ. 법관의 제척·기피·회피	47
1. 의 의	47
2. 제 척	47
3. 기 피	49
4. 회 피	54
제2항 당사자	54
Ⅰ. 당사자의 의의	54
1. 형식적 당사자개념	54
2. 당사자의 호칭	55
3. 당사자는 민사소송의 주역-당사자권	55
4. 두 당사자 대립의 구조	55
Ⅱ. 당사자의 확정	56
1. 의 의	56
2. 당사자확정의 기준	56
3. 당사자확정이 문제되는 영역	57

What represents a textbook is the table of contents. The table of contents represents the system of the textbook. In this way, the textbook arranges all the knowledge in the world in an orderly manner in one book. Whenever a question is asked, it is answered by extracting the necessary parts from the whole system of the textbook and organizing them again. Therefore, a well-crafted textbook can answer "everything," at least as far as your field is concerned. It is infinitely answerable. This is the textbook, this is the system, and this is the K-engine.

(2) Knowledge Unit

What a textbook contains is knowledge. Although it is not absolute, knowledge is usually contained in one sentence, one sentence is assigned to one piece of knowledge, and one title is attached to one piece of knowledge, thus, the pair of subject and content is one unit of knowledge.

The K-Engine reorganizes all the knowledge in the textbook in this way. This is because only then the computer can consume and digest the knowledge. Once the computer has ingested and digested the knowledge, it becomes a human being. It becomes a lawyer, it becomes a doctor, it becomes... It becomes an expert in every field in the world. I've mentioned this before, but after consuming about 10,000 pages of knowledge in each field, there will be no more experts left in the world. They will only take supporting roles. 10,000 pages is what 100 people can input in 100 days. With 10 people, that's 1,000 days which is three years. Eventually, in about two years, all the knowledge of the world will be sucked into the computer. Heaven on earth is not far away.

A knowledge unit consists of a sentence that represents the knowledge itself and various information that the knowledge has in relation to other knowledge. In other words, title, content, and linked information are the basic elements that make up a knowledge unit. The four components of a knowledge unit are completed by adding location information corresponding to the systematic information of the textbook itself. One example is as follows. First, imagine the following scenario. a client asks:

"Whether I can change the residency if an organization is approved as a corporation."

If the client asks K-Engine a question with the keyword or sentence, this would typically result in the following 17 knowledge items.

5. Title

(1) Restrict change to resident or non-resident

6. Content

(1) In case an organization is approved as a corporation, it may not change the status to resident or non-resident in the taxation period including the approval date and until the business year including the day after three years from the end of the taxation period; unless the approval is canceled because of failure to meet the requirements for approval by the Head of Tax Office.

7. Location information

(1) Location information of the entire system

i. Corporation tax return > A. Tax obligations of corporation > iii. Tax obligations by corporation type > (3) Tax obligations of non-profit corporations > 2) Types of non-profit corporations > ③ Organizations deemed as corporations under §13 of the Framework Act on National Taxes > B. Organizations deemed as corporations under approval by the Head of the Tax Office

(2) Location information ID

i. 17

8. Linked information

(1) e3<n>,r13<same as the above>

As explained above, we don't need to explain the title and content, but only the location information and linked information which is a tag.

(3) Location Information

Location information is the coordinate information where the knowledge is located in the entire system. It can be expressed in two forms; one is a way to indicate all the relationships with the entire system, and the other is the location information that identifies the location.

Location information is further divided into three types:

i. Table of Contents of Parent

Systematic knowledge is in the form of a flowchart, and the flowchart is based on a parent-child relationship. Therefore, any knowledge unit that exists on the flowchart can identify its position in the system by knowing who its parents are.

ii. Level information

This is actually information that is determined naturally in the process of identifying parent information. All knowledge in the diagram is organized into parent and child units, and in this case, the child unit is one level lower than its immediate parent unit. If 1. is followed by 1), then 1) is one level below 1. In the printout, it will appear as a single tap to the right.

iii. Own order information on the same level

This is the order of information in the relationship between other knowledge units of the same level if they are equally located at the level where you are located. Thus, it is the sequence number in the sibling relationship. Therefore, if you are listed through the table of contents such as 1, 2, 3, ... or A, B, C, ... on the same level, if you are B, then B will be your order information.

(4) Linked information

Linked information is information that indicates the connection between the knowledge and other knowledge. In the above example, there are two tags: e3 and r13. (For reference, the letters in <> mean a conjunction when connected to knowledge 13, and 'n' means no conjunction is required.)

e3 is an essential companion to knowledge 3. Knowledge 3 is as follows:

- 4. Title
 - A. An organization deemed as a corporation by the approval of the Head of the Tax Office
- 5. Content
 - A. Certain types of corporations that do not distribute profits to their members are viewed as a corporation and subject to taxation.
- 6. Location Information
 - A. Location information in the entire system
 - i. Omitted
 - B. Location information ID
 - i. 15

r13 is a Recommended companion knowledge. Knowledge 13 is as follows:

4. Title

(1) Requirements

5. Content

(1) A corporation, foundation, or other entity must have regulations governing the organization and operation of the corporation, foundation, or other entity and appoint a representative or manager.

(2) A corporation, foundation, or other organization must independently own and manage its revenues and property on its own account and in its own name.

(3) Not distribute the profits of the corporation, foundation, or other organization to its members.

3. Location Information

A. Location information in the entire system

i. Omitted

B. Location information ID

ii. 13

Generally, a single knowledge is divided into a single sentence, but in some cases, as shown above, three very closely related knowledge can be combined into a single knowledge. The result of combining these three pieces of knowledge is the following.

1. An organization deemed to be a corporation by the approval of the Head of the Tax Office

Certain types of organizations that do not distribute their profits to their members are deemed to be corporations and are subject to taxation.

2. Requirements

(1) A corporation, foundation, or other entity must have regulations governing the organization and operation of the corporation, foundation, or other entity and appoint a representative or manager.

(2) A corporation, foundation, or other organization must independently own and manage its revenues and property on its own account and in its own name.

(3) Not distribute the profits of the corporation, foundation, or other organization to its members.

3. Restrictions on changing to resident or non-resident status

In case an organization is approved as a corporation, it may not change the status to resident or non-resident in the taxation period including the approval date and until the business year including the day after three years from the end of the taxation period; unless the approval is cancelled because of failure to meet the requirements for approval by the Head of Tax Office.

This final result is what the K-Engine would display if the original 17th knowledge was searched for.¹⁰ But it doesn't end there. The result above is,

"Whether I can change the residency if an organization is approved as a corporation."

but in reality, it is common for dozens of knowledge to be extracted. In other words, the above example is just the output based on knowledge 17, which was the most accurate.

K-Engine usually generates output based on the top three or four search results, thus, if the target knowledge with a rank score of 100 is #17, the target knowledge with a rank score of 85 is #34, and the target knowledge with a rank score of 77 is #5, and their connection information is as follows:

- 7th knowledge
 - rank : 100
 - linked information (tag) : e3<n>,r13< same as the above >
- 34th knowledge
 - rank : 85
 - linked information (tag) : e22<n>,r3<n>, r19<n>
- 5th knowledge
 - rank : 77
 - linked information (tag) : r27<n>,r35<n>, r41<n>

The final extracted knowledge data is as follows:

e3, r3, **5**, r13, **17**, r19, e22, r27, 34, r35, r41, **77**

The thick numbers are the original extracted knowledge data, and the others are the knowledge data extracted together through linked information based on them. Among them, e3 and r3 are duplicates, so they only need to be output once. Therefore, for a

¹⁰ The above result shows that the conjunction "same as the above" has been added.

total of 10 pieces of knowledge, the final result set of IDs is:

3, 5, 13, 17, 19, 22, 27, 34, 35, 41

If we were to extract and expose only the required companion knowledge, only the following 5 knowledge will be exposed:

3, 5, 17, 22, 77

All of this knowledge has its own systematic location information. As mentioned earlier, the location information is divided into the following three types:

- Table of contents of parent
- Level information
- Own order information on the same level

Based on this information, K-Engine assigns a new table of contents to the extracted knowledge and outputs the final result in an orderly manner.

3. Diversity of Tag Criteria

The tag symbols representing linkage information (e3<n>,r13<as above> - linkage information for knowledge No.7) were manually extracted based on the target knowledge data (i.e., knowledge No.7) and the knowledge that is linked to it. In other words, if only the target knowledge data (No. 7) is extracted, it may be unclear in what context this knowledge appears, and even the meaning of the knowledge itself may be difficult to understand. Therefore, we need supplementary knowledge data that can supplement these deficiencies. In other words, the extraction criteria for companion knowledge data in this case is "semantic self-completeness."

The problem is that "semantic self-completeness" is not the only criterion for extracting companion knowledge data, but one of many. As mentioned above, it is possible to extract other criteria other than semantic self-completeness, such as "companion knowledge related to real-world cases based on the real-world cases." In this way, you can organize your tags by first collecting a large number of real-world examples and then extracting the companion knowledge associated with each one. This is a kind of "multiple perspectives of answers" to "multiple perspectives of questions." And those perspectives can be combined in a multi-dimensional form.

In other words, if we take the "self-completeness of semantics" of the previous most basic case as our basis, then we can use:¹¹

1. Semantic self-completeness
2. Real-world relevance

these above companion knowledge in two dimensions.

To put it more simply, a question about food can be asked from a nutrient perspective, a diet perspective, a price perspective, and many other perspectives. The answers to these questions will vary depending on the result set of the accompanying knowledge. And the most basic of these perspectives is the self-completeness of meaning.

We can also extract companion knowledge in two dimensions. In this way, not only one-dimensional extraction but also multi-dimensional extraction will significantly exceed the level of the human brain. This is because the capacity and speed of the human brain are overwhelming, even for one-dimensional companion knowledge extraction. However, computers can reach the second dimension, third dimension, fourth dimension... and even infinite dimensions within their capacity, which can be

¹¹ In most cases, this is indeed the case. This is because semantic self-completeness is the most basic connection information extraction criterion.

indeed seen as the appearance of God.

4. Human Structure of Thinking and K-Engine

A "system" can contain all the knowledge in the world. It was enough for now. Until now, all knowledge could only be contained in a tangible object called a book, and the work could only be done to organize it, with no further processing. This is how knowledge has been organized, transmitted, understood, and taught for thousands of years in the form of "books".

When a problem arises, the first thing we do is turn to a book. Of course, those who have the book well digested and organized in their heads, such as lawyers, don't need to look through a book, they can use their brain book. This brain book is very fast, just like a computer. When asked a question, the lawyer responds immediately. This is the result of the information stored in the brain communicating with each other at high speed through the neural network.

In this scene, we can assume that somewhere in the lawyer's brain, linked information between knowledge is stored. This linked information interacts to create a beautifully structured answer that flows out of his mouth. This is the implementation of the K-Engine.

As you can see from the above, so far, we have only created a tangible "book" offline, and the linked information between the knowledge in the books has been stored with great effort in the neurons of the human brain. In other words, the two have been stored separately.

The K-engine is the integration of these two separate elements. The computer stores all of these elements on a single hard disk, which means that you don't have to make any effort to write a book anymore. Anyone can become a lawyer without studying. The same goes for doctors and pharmacists. All the expertise in the world is yours without having to "study" it. Furthermore, general information obtained through Google is also systematized. Thus, the area of obtaining daily knowledge will also come under the control of the K-Engine.

Once the human connection information is entered, all questions can be answered based on the systematic nature of the textbook. At this time, the manual work is not done by me, but by others. I don't have to do anything. I can receive perfect grades without studying. There is no reason to go to school.

5. Expansion of K-Engine

At the time of writing this White Paper, only one section of the 700-page corporate tax among tax law is being inputted. There are two directions for the expansion:

(1) Vertical Expansion

Personally, as I mentioned earlier, I expect to eventually have 100,000 pages of data entry per field. This will make you better than a lawyer with decades or centuries of experience. No matter how much detail there is, no matter how much complexity and depth there is, K-Engine will pick it up like a tweezer. Lawyers and other experts have to review the data, but K-Enterprise provides answers instantly. With just a single input, the entire human race becomes lawyers and doctors. However, updates are required. Therefore, existing lawyers and doctors can continue their roles as personnel in charge of updates.

The principle of the K-Engine may seem simple, but that simpleness is enough to contain all the knowledge in the world. The simpler it is, the more accurate it is. The simpler it is, the less risk of error. The simpler it is, the more infinite it is. The simpler it is, the more highly it can evolve. The evolution of K-Engine is infinite.

In my opinion, it is unlikely that we will see any more K-Engines. Even if they do, there is no need for them. It's like saying that there's no need to develop a new OS when Windows exists today. In other words, just as Windows is the end of the PC OS, K-Engine is the end of K-Engine. This is because K-Engine is so simple that it can't be simplified any further. It seems impossible that there will be any more patents.

In conclusion, K-Engine is the definitive version of conversational AI, and there is no more. The future growth of conversational AI will come from within K-Engine, not outside of K-Engine.

(2) Horizontal Expansion

I plan to establish 1,000 corporations with K-Engine as the holding company and list them all on NASDAQ or KOSDAQ. I plan to grant licenses individually to encourage competition based on license renewal. If a licensed company becomes lazy, it will not be renewed and will be licensed to another company that wants to compete.

The input of basic data for K-engine is believed to be around 1 billion won per field. If there are 1,000 companies, it will cost about 1 trillion won. I don't plan to do this myself, but at least I plan to exercise control based on licensing.

As time goes by, the fields will expand and be subdivided. It is expected to be subdivided into about 100,000 fields. From the point of view of the Korean government, the country where I was born, if only 10 people are employed in each field on average, a total of 1,000,000 jobs will be created, and Korea's GDP per capita will be the highest in the world. It is also speculated that even the national GDP of Korea could become the highest.

(3) Expanding Connecting Logic

This is why there's so much more development to come for K-Engine. It is based on the tagging criteria as mentioned earlier, namely based on the diversification of connection logic and its high dimensionality. Hereby, K-Engine will expand horizontally to all fields, add vertical depth to each field, and finally, ultra-sophisticate the human knowledge system through variation of various perspectives. In this way, artificial humans will become true humans, and current humans will be reduced to the level of artificial humans. I want my children to be better than me. Similarly, I want K-Engine to be more than human, even though it is made by human hands.

One last question. I think I've proven enough that K-Enterprise is the ultimate protocol for artificial intelligence. Any objections?

6. Comparison with GPT

Comparison	GPT	K ENGINE
Operation base	Language	Knowledge
Basic data	Information on the web	Textbook knowledge
Database capacity	Big-data based	Small-data based ¹²
Possibility of inheriting error from source data	YES	NO ¹³
Possibility of error in the combining process of knowledge	YES	NO
Answers for Expert Knowledge	WEAK	STRONG
Answers for General Knowledge	STRONG	WEAK in short term STRONG in long term ¹⁴
Depth of Answer	Shallow	Deep
Copyright issue	POSSIBLE	NONE ¹⁵
Server specification	High specification required	Does not require high specification
Hard disk specification	Ultra-large specification required	100 megabytes per field is sufficient
Power consumption	Huge	Same as a general server (e.g. Google server, etc.) ¹⁶

¹² The amount of data required to produce a lawyer-quality answer in any one field is about 10 megabytes (10,000 pages).

¹³ Of course, if there are errors in the production of the textbook itself, the K-Engine could also make errors. Of course, this is only theoretically possible.

¹⁴ In the long term, all general knowledge will be included in the K Engine. At that point, K-Engine will surpass GPT in the field of general knowledge.

¹⁵ GPT inevitably encounters copyright issues in the process of sweeping up all the knowledge on the Internet, but K-Engine is completely free from copyright issues because it creates its own data (textbooks).

¹⁶ It will use much less power than Google. This is because Google is also based on big data. K- Engine only competes with logic based on small data that has been essentialized, so the amount of power required is extremely small.

7. Related Patents

- Korean Patent Application No.10-2023-0053098 : Connection Knowledge Generating Method and System Through Knowledge Crossing and Knowledge Connection)
- Korean Patent Application No.10-2023-0108659 : Method of generating combined knowledge through extraction of accompanying knowledge)

