**넷플릭스 데이터를 통한 리뷰 데이터 등록 서비스 제안**

Proposal of review data registration service through Netflix data

**요 약**

**넷플릭스는 전세계적으로 OTT 서비스의 선두주자격인 플랫폼이다. 하지만 서비스 사용자들의 행동이나 구매에 영향을 미칠 수 있는 온라인 리뷰를 남길 수 있는 기능이나 서비스는 제공되지 않고 있다. 따라서 기업을 지속적으로 성장시키기 위해 사용자들이 리뷰를 등록하고, 다른 사용자들의 리뷰를 볼 수 있는 서비스가 필요하다. 해당 연구는 넷플릭스 서비스 사용자들이 서비스 환경 내에서 리뷰를 등록 및 확인할 수 있는 서비스를 위한 데이터베이스 구축을 제안한다. 먼저 넷플릭스에 등록된 콘텐츠들에 관한 정보와 리뷰를 4flix에서 수집한다. 수집한 데이터를 데이터베이스에 등록하기 위해 데이터베이스를 정의한다. 데이터베이스의 정의는 Entity set, Relation, Attribute를 정의하고, 각 Attribute의 primary key와 테이블간의 연동을 위한 foreign key를 정의하는 것으로 이루어진다. 수집한 데이터셋을 데이터베이스에 정의된대로 입력하였다. 데이터베이스에 입력된 데이터들을 바탕으로 데이터베이스간 연동이 잘 되었는지 확인하였다. 연동을 확인한 방법으로는 특정 리뷰에 대해 조회한 결과 해당 리뷰에 작성된 댓글이 나오는 것을 확인하였고, 영화 제목을 입력하면 해당 영화에 대해 남겨진 리뷰와 평점들을 모두 볼 수 있음을 확인하였다. 또한 영화 제목을 입력하면 서비스 사용자들이 남긴 해당 영화에 대한 평균 평점도 볼 수 있었다. 해당 연구가 제시한 넷플릭스 컨텐츠에 대한 리뷰 등록 서비스와 데이터베이스를 이용하면 서비스 사용자들의 구매행위에 영향을 줄 수 있고, 기업에게 시장 경쟁우위를 가져다 줄 수 있을 것이다.**

**Abstract**

Netflix is a globally qualified platform for OTT services. However, there are no features or services available to leave online reviews that can affect the behavior or purchase of service users. Therefore, in order to continue to grow the company, there is a need for services that allow users to register reviews and view reviews from other users. The study proposes to build a database for services that allow Netflix service users to register and check reviews within the service environment. First, 4flix collects information and reviews about the content registered on Netflix. It defines a database to publish the collected data to the database. The definition of a database consists of defining the entity set, relationship, and attribute, and the foreign key for interworking between the primary key and the table of each attribute. The collected datasets are entered as defined in the database. Based on the data entered into the database, it was checked that the inter-database connection was successful. As a method of checking the link, we checked for a specific review and found that the comments written in the review came out, and if we entered the title of the movie, we could see all the reviews and ratings left for the movie. In addition, if you enter the title of the movie, you can see the average rating of the movie left by the service users. Using the review registration service and database for Netflix content suggested by the study could affect service users' purchasing behavior and give companies a competitive advantage in the market.

Keywords: netflix, database, review service

1. **Introduction**

넷플릭스는 OTT서비스의 대표적인 선두주자이다. 미국과 캐나다에서 2300만명 이상의 구독자를 유치하고 있다 [1]. 이 때 온라인 리뷰는 서비스 사용자들이 각자의 소비 경험을 바탕으로 디테일하고 믿을만한 정보를 얻는 정보의 원천으로 알려져있다 [2][3]. 그리고 이러한 온라인 리뷰들을 통해 서비스 사용자들은 상품 구매시 불확실성을 줄이고자 하는 경향을 보인다 [4].

하지만 넷플릭스 사용자들은 넷플릭스 사용환경에서 특정 컨텐츠에 대한 다른 이용자들의 리뷰를 볼 수 없고, 자신의 리뷰 또한 남길 수 없는 상황이다. 서비스 사용자들이 넷플릭스의 컨텐츠와 관련해 리뷰를 남기려면 관련된 웹사이트에 접속해 회원가입을 한 뒤, 컨텐츠를 검색 및 선택하고 자신의 리뷰를 남기는 방법밖엔 없다. 따라서 넷플릭스 사용환경에서 특정 컨텐츠에 대한 리뷰를 직접 볼 수 없고, 리뷰를 작성하기 쉽지 않기 때문에 온라인 리뷰 데이터베이스의 활성화도 쉽지 않을 것이다. 따라서 본 연구에선 넷플릭스 컨텐츠에 대한 리뷰를 남길 수 있게 만들어진 웹사이트에서 데이터 크롤링을 진행하고, 뒤에 설명할 E/R diagram을 바탕으로 관계형 데이터베이스를 구축한 뒤, 이를 바탕으로 사이트에 따로 들어가지 않아도 넷플릭스 사용환경에서 바로 특정 컨텐츠에 대한 다른 이용자들의 리뷰를 볼 수 있고, 자신의 리뷰도 남길 수 있는 리뷰 데이터베이스를 제안하고자 한다.

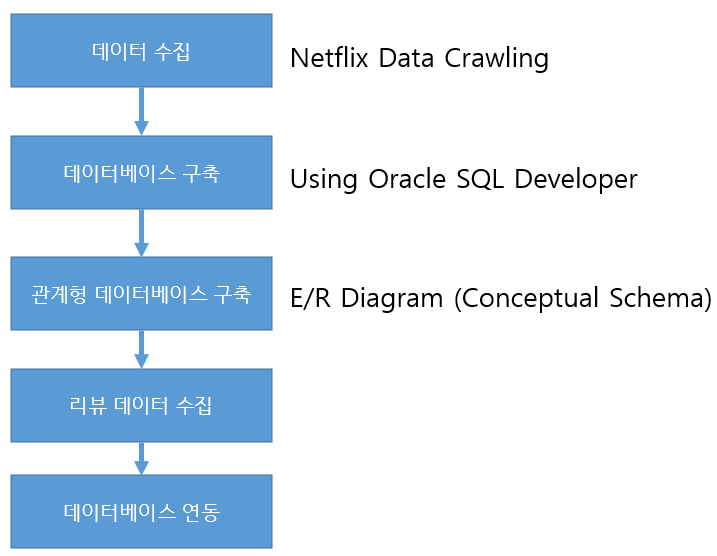
1. **Related Work**

온라인 리뷰 데이터베이스는 제품 및 서비스 품질에 관한 정보의 원천으로서 전문가의 의견과 사회적 학습을 보완하고 있다. 이러한 데이터베이스는 사용자가 남긴 리뷰가 많은 사람들에게 접근할 수 있도록 허용해준다. 관련된 연구로는 이베이 셀러 레이팅이 이베이 입찰자들의 행동에 어떤 영향을 주는지에 대한 연구가 있다 [5][6]. 해당 연구에선 이베이 셀러의 평균 레이팅을 알고 있고, 셀러에 대한 다른 정보가 없을때 온라인 리뷰가 입찰자의 행동에 영향을 미친다는 것을 보여줬다. 온라인 리뷰가 책의 판매 순위에 미치는 영향을 조사한 연구도 있다. Chevalier et al (2006).은 Amazon.com과 Barnesandnoble.com(bn.com)에서 고객의 리뷰가 책의 판매 순위에 미치는 영향을 조사했다. 해당 연구에선 bn.com보다 Amazon.com에서 더 좋은 평가를 받은 책이 bn.com에서 더 잘 팔리는 경향이 있다는 것을 발견했다 [7].

Boot (2017)은 서적에 대한 연구를 용이하게 하기위해 온라인 서점의 서적에 대한 반응뿐만 아니라 네덜란드어로 쓰인 리뷰들을 방대하게 가지고 있는 리뷰 웹사이트에서 데이터를 가져와 Online Dutch Book Response (ODBR) 데이터베이스를 구축하였다. 해당 연구의 목적은 온라인에서의 서적 반응에 대한 연구를 용이하게 해주는 데이터베이스 구축이기 때문에 그에 관련한 E/R diagram을 구축하였고 book lists, expert reviews, blog post 등과 같은 entity set을 정의하였다. 그리고 이를 바탕으로 146,800권의 서적과 58,000개의 사용자 계정, 628,800개의 서적 리뷰 등을 포함하는 데이터베이스를 구축하였다. 구축한 데이터베이스를 이용해 진행할 수 있는 연구를 보여주기 위해 데이터베이스의 리뷰를 이용해 장르별로 클러스터링을 진행하였고, 데이터베이스를 이용하여 진행할 수 있는 향후 연구들의 가능성을 제시하였다 [8].

1. **Methods**

**3.1 Framework**



**Figure 1. Framework**

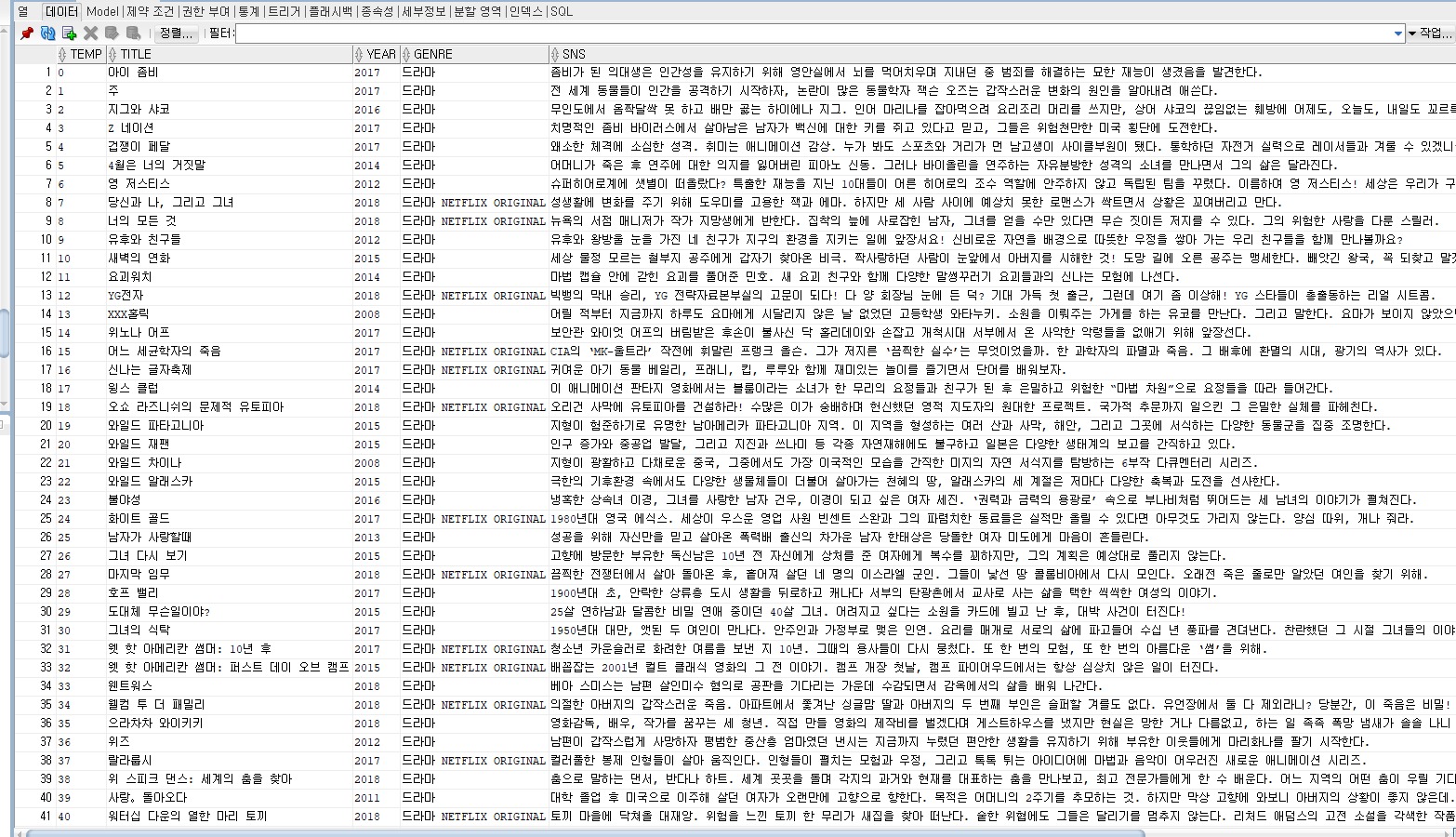
본 연구에서의 Framework는 다음과 같이 구성하였다. 먼저, Netflix 데이터를 수집하기 위해 Netflix에서 상영하는 영화의 리뷰를 구축하고 있는 4flix 데이터를 이용하여 수집하였다. 수집한 데이터는 Oracle SQL Developer를 통해 데이터베이스를 구축하였다. 새로운 관계를 구성하기 위해 E/R Diagram을 통해 구축할 데이터베이스를 확정하고 Oracle SQL Developer를 통해 새로운 개념적 스키마들을 추가하였다. 그리고, 리뷰 데이터를 수집하게 되면, 수집 데이터들이 데이터베이스에 연동할 수 있도록 데이터 구조를 재정의하고자 한다.

**3.2 Data Collection**

먼저, 컨텐츠 수출 및 수입 시 컨텐츠의 제목이 달라지는 문제가 존재하기 때문에 한국 넷플릭스에 적용하기 위해 컨텐츠에 대한 데이터를 가지고 있는 4Flix에서 데이터 크롤링 작업을 실시하였다. 크롤링 작업을 하기위해 파이썬 버전 3.7.9를 사용하였다. 수집한 데이터로는 title(제목), year(연도), genre(장르), synopsis(줄거리)를 수집하였다. 총 수집한 데이터는 4,417개이다.

**3.3 Database Structured**

데이터베이스 구축을 위해 Oracle SQL Developer를 사용하였다. 수집한 데이터를 연동하기 위해 따로 Movie Model Table을 구축하여 다음과 같이 구성하였다. Movie Model Table의 Attribute에는 Title, Year, Genre, Synopsis로 구성하였다. 또한, 한글로 구성된 데이터이므로, 환경 변수는 ‘UTF-8’로 설정하였다.

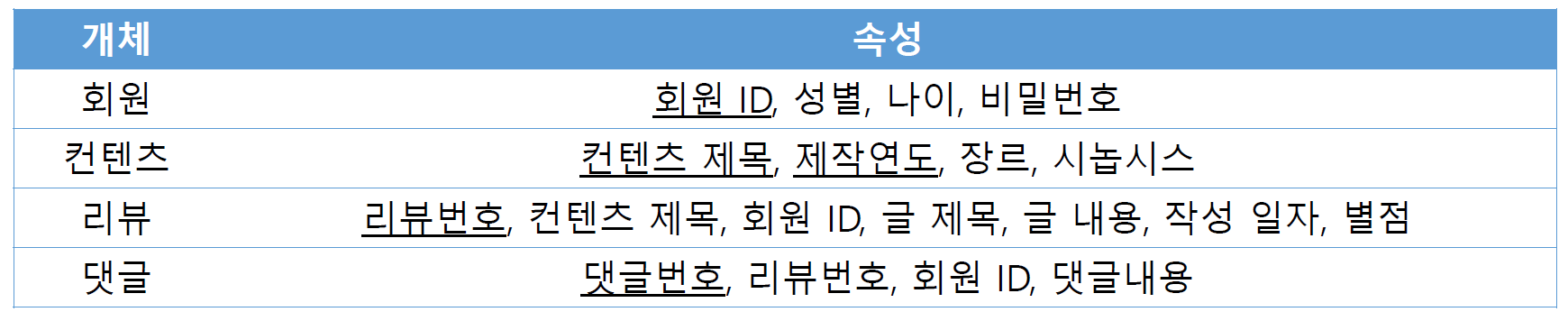


**Figure 2. Imported data in SQL Developer**

**3.4 Relational Database Structured**

**3.4.1 Conceptual Design**

먼저, 관계형 데이터베이스를 구축하기 위해 개념적 설계부터 작성하였다. 총 4가지의 Entity Set으로 구성하였다. 구성된 Entity Set으로는 회원, 컨텐츠, 리뷰, 댓글로 확인할 수 있다. 이미, 데이터 구축 때, Movie Model Table이 컨텐츠 Entity Set이라고 볼 수 있다. 또한, 각 Entity Set에서의 Primary Key를 지정하였다. 이때, 컨텐츠 Entity Set에서는 컨텐츠 제목과 제작연도가 같이 묶어서 Primary Key로 지정하였다. 제목이 같으면서도 리믹스 된 영화들이 나타나면서 구분을 해주기 위해 제작연도까지 포함해서 Primary Key로 설정하였다.



**Figure 3. Database concept**

**3.4.2 Relationships of Entity Set**

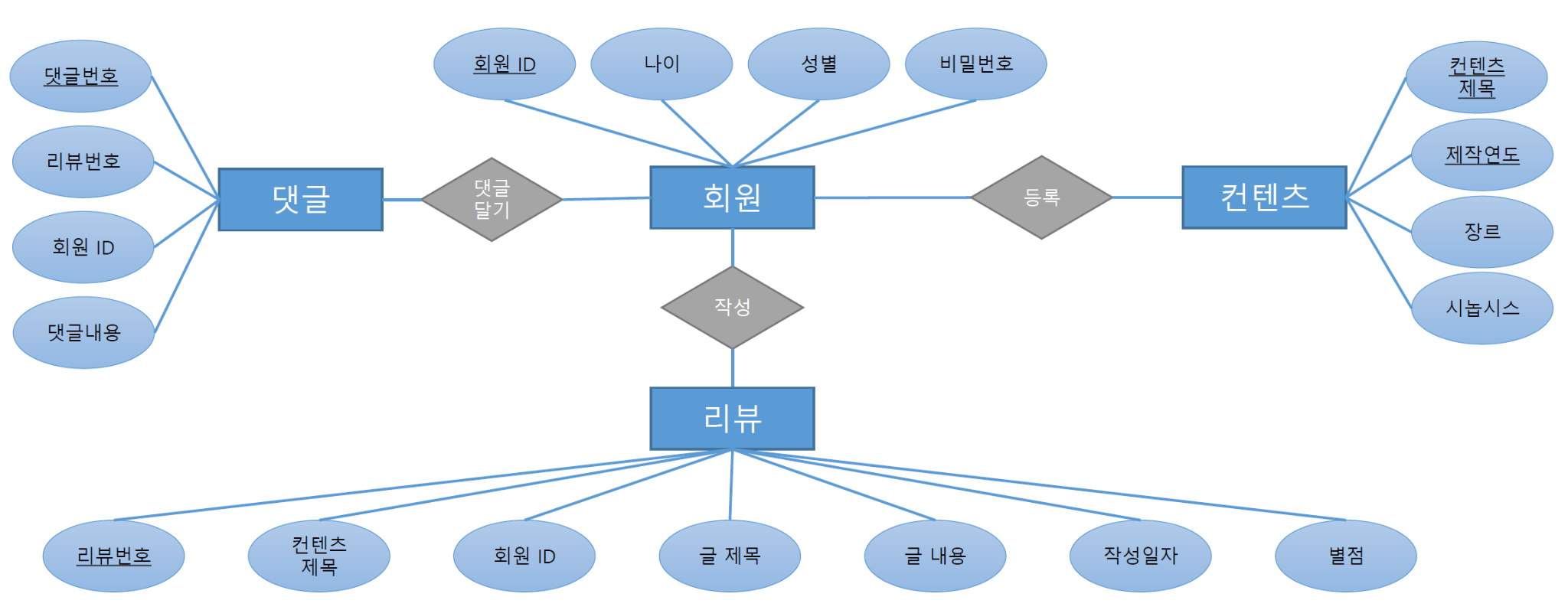
4가지의 Entity Set들을 연결시키기 위한 Relationship을 정의해야한다. 한 회원이 여러 영화를 볼 수 있기 때문에, 회원과 컨텐츠는 1:N의 관계로 설정하였다. 또한, 한 회원이 여러 영화의 리뷰를 남길 수 있기 때문에, 회원과 리뷰는 1:N의 관계로 설정하였다. 한 회원이 여러 리뷰에 대하여 댓글을 남길 수 있으므로, 회원과 댓글은 1:N의 관계로 설정하였다.



**Figure 4. Relation definition**

**3.4.3 Conceptual Schema**

개념적 설계 및 개체 간의 관계를 설정한 후 이를 바탕으로 개념적 스키마를 구축하였다. E/R Diagram을 확인하면 3가지의 Entity Set을 추가적으로 구축하고자 한다. 먼저, 회원 Entity Set에서는 회원ID가 Primary Key이고, Attribute로 나이, 성별, 비밀번호가 포함된다. 리뷰 Entity Set에서는 리뷰번호가 Primary Key로 식별할 수 있는 key가 된다. Attribute로 컨텐츠 제목, 회원ID, 글 제목, 글 내용, 작성일자, 별점으로 구성한다. 댓글 Entity Set에서는 댓글번호라는 식별할 수 있는 Primary Key를 지정한다. Attribute로 리뷰 Entity Set과 연결하는 리뷰번호, 회원 Entity Set과 연결하는 회원ID, 댓글 내용으로 구성한다. 댓글 Entity Set에서는 리뷰 번호와 회원ID는 Foreign Key (외래키)인 것을 확인할 수 있다.



**Figure 5. E/R Diagram of the Database**

**3.5 Review Data Collection**

영화 관련 리뷰 데이터를 구축할 때, E/R Diagram으로 통해 구축한 데이터베이스가 잘 작동되는지 확인할 예정이다. 리뷰 데이터가 잘 들어가는지 확인한다.

**3.6 Database Interworking & Function of DB**

데이터베이스의 연동을 확인하고 추후 넷플릭스 관련 데이터도 분석해볼 예정이다.

1. **Results**

**4.1 Entity Set Configuration**

**4.1.1 회원 Entity Set**

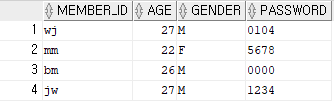


Figure . 회원 Entity Set

회원 Entity Set은 그림6과 같이 MEMBER\_ID, AGE, GENDER, PASSWORD로 Attribute로 구성했다. 그리고, MEMBER\_ID를 Primary Key로 지정하였다.

**4.1.2 컨텐츠 Entity Set**

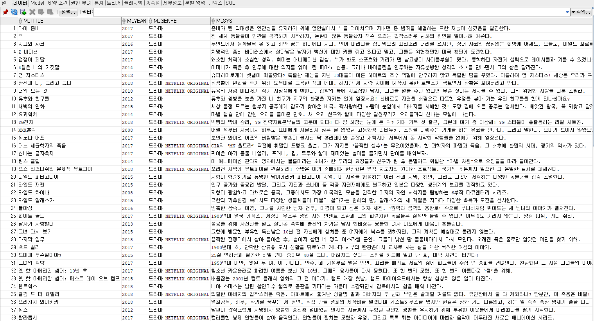


Figure . 컨텐츠 Entity Set

컨텐츠 Entity Set은 그림[]과 같이 M\_ID, M\_TITLE, M\_YEAR, M\_GENRE, M\_SYS로 Attribute로 구성했다. 그리고, M\_ID, M\_TITLE를 Primary Key로 지정하였다.

**4.1.3 리뷰 Entity Set**

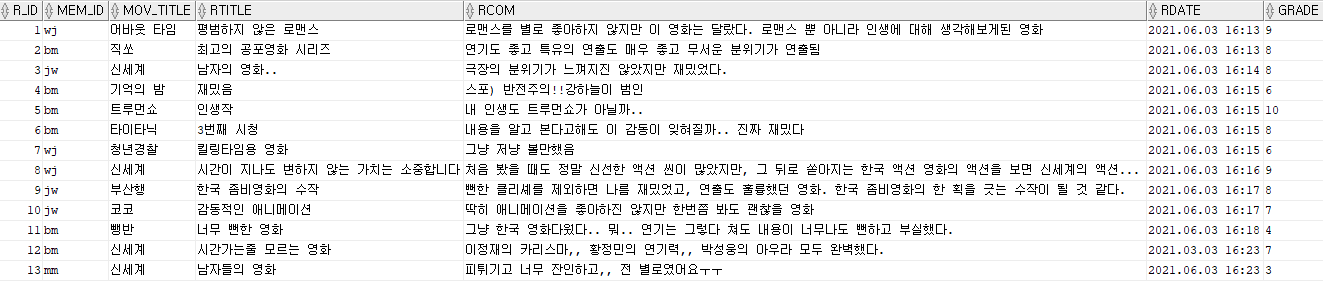


Figure . 리뷰 Entity Set

리뷰 Entity Set은 그림[]과 같이 R\_ID, MEM\_ID, MOV\_TITLE, RTITLE, RCOM, RDATE, GRADE로 구성했고, R\_ID를 Primary Key로 지정했다.

**4.1.4 댓글 Entity Set**

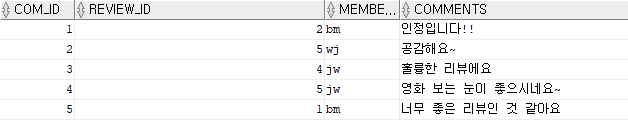


Figure . 댓글 Entity Set

댓글 Entity Set은 그림[]과 같이 COM\_ID, REVIEW\_ID, MEMBER\_IT, COMMENTS로 Attribute로 구성했다. 그리고, COM\_ID를 Primary Key로 지정하였다.

이와 같이 Oracle SQL Developer를 통해 각 Entity Set을 구성하였다.

**4.2 Configure Entity Set Relationships**

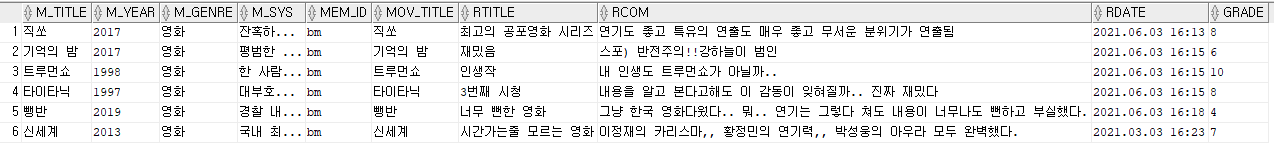
****

Figure 10. 회원, 컨텐츠 간의 1:N의 관계 및 회원, 리뷰 간의 1:N의 관계

그림[]은 SQL문을 실행한 결과로, 회원과 컨텐츠 간의 1:N 관계, 그리고 회원과 리뷰간의 1:N 관계를 모두 보여준다. 한명의 회원이 여러 컨텐츠에 접근했고 또한 한명의 회원이 여러개의 리뷰를 남겼음을 볼 수 있다.



Figure 11. 회원과 댓글 간의 1:N의 관계

그림[]은 SQL문을 실행한 결과로, 회원과 댓글간의 1:N 관계를 보여준다. 한명의 회원이 여러 리뷰에 대한 댓글을 작성했음을 볼 수 있다.

**4.3 Apply Review Data and Comment Data**

**4.3.1 Apply Review Data**



Figure . Review Data Code

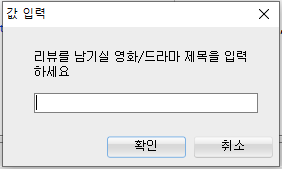


Figure . Write Movie Title

SQL문의 INSERT문과 ACCEPT PROMPT 구문을 사용하여 R\_ID는 입력받을 때마다 1씩 늘어나게 저장하고, RDATE는 사용자가 입력한 날짜와 시간을 나타낸다. 그리고 나머지 사용자에게 직접 입력받은 값, 자신의 회원 id, 리뷰를 남기고자 하는 컨텐츠 제목, 리뷰제목, 리뷰 내용, 평점을 모두 파악하여 리뷰 테이블에 저장한다.

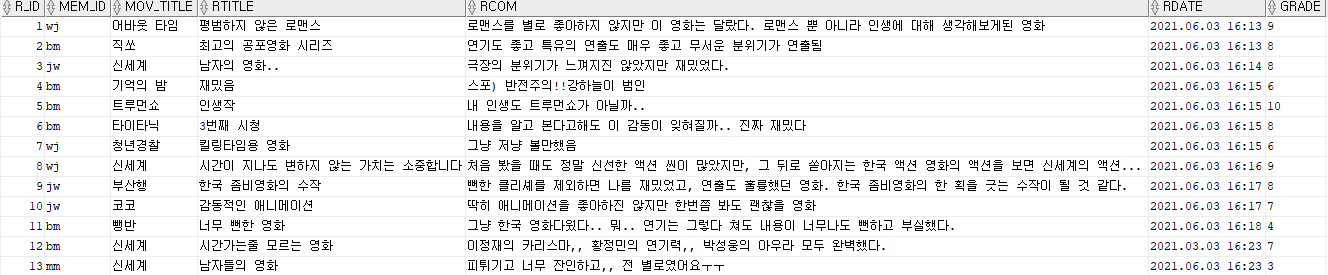


Figure . Apply Review Data

이 과정을 반복 실행하였고, 수천개의 데이터를 구성함에는 무리가 있어, 본 연구에 대한 설명에 필요한 13개의 리뷰데이터를 작성하였다.

**4.3.2 Apply Comment Data**



Figure . Comment Data Code

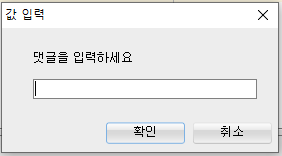


Figure . Write Comment

댓글의 입력 과정도 리뷰와 같다. SQL문의 INSERT문과 ACCEPT PROMPT 구문을 사용했고, 특정 리뷰에 대한 댓글을 작성했다.

Figure . Apply Comment Data

연구에 대한 설명에 활용할 댓글 데이터는 그림[]의 5개와 같다.

**4.4 Database Interworking**

**4.4.1 Interlink reviews and comments**

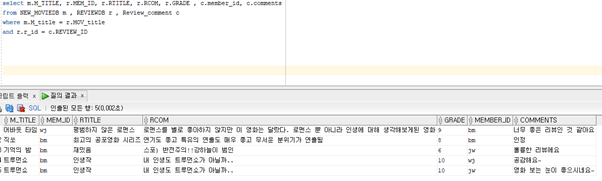


Figure . Interlink reviews and comments

리뷰와 댓글의 연동에 대해 SQL 질의문을 작성했다. 그 결과는 그림[]과 같다. 특정 리뷰에 대해 조회한 결과 해당 리뷰에 작성된 댓글이 나오는 것을 볼 수 있다.

구성하고자 하는 E/R Diagram처럼 각 Entity Set들을 생성하고 각 Entity Set을 Relationship한 모습을 확인하였다. 또한, 실제로 적용할 경우 리뷰 및 댓글이 입력되는 것을 확인할 수 있다. 이처럼 실시간 리뷰 및 댓글을 확인할 수 있으면 효율적인 데이터베이스를 구축할 수 있을것으로 보인다.

**4.5 Function of DB**

**4.5.1 View all reviews of the movie**

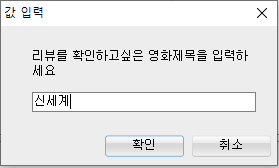


Figure . Select movie

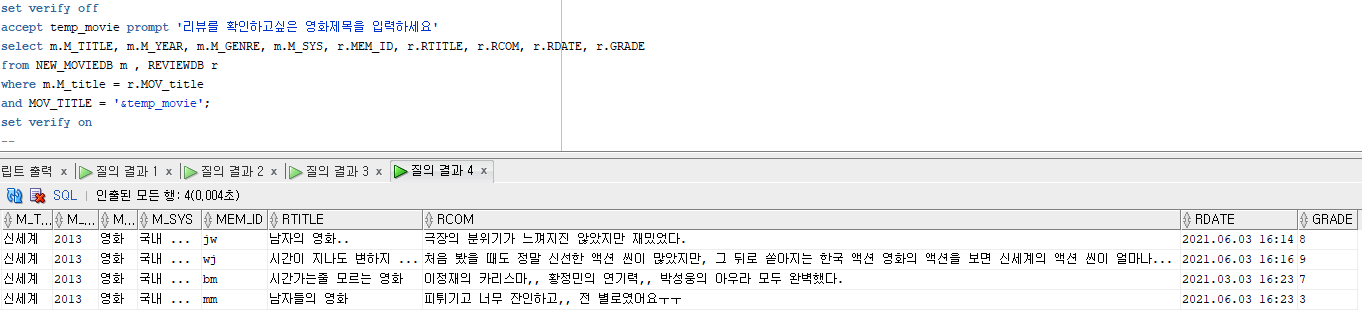


Figure . Results of all reviews of selected movie

본 연구에서 설계된 DB의 기능은 다음과 같다. 우선 그림[]과 같이 영화 제목을 입력하면 해당 영화에 대해 남겨진 리뷰와 평점들을 모두 볼 수 있다.

**4.5.2 Check the average rating of the movie**

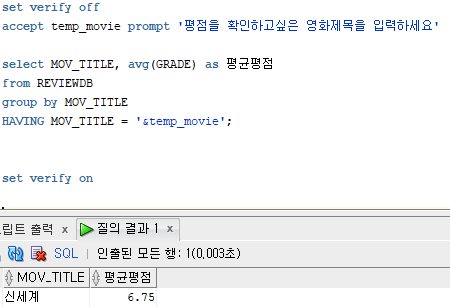


Figure . Average rating of the movie

두번째 기능은 평균평점 확인이다. 사용자가 영화 제목을 입력하면 사용자들이 남긴 해당 영화에 대한 평균 평점을 볼 수 있다.

1. **Conclusion**

본 연구에선 넷플릭스에 등록된 콘텐츠에 대한 제목, 시놉시스, 리뷰 등을 수집하여 기존 넷플릭스 사용자들이 서로의 리뷰를 확인할 수 있고, 자신의 리뷰를 등록도 할 수 있는 데이터베이스를 구축하여 새로운 리뷰 등록 서비스를 제안하였다. 데이터의 수집은 넷플릭스의 컨텐츠 목록을 보유하고 있는 4flix에서 진행했다. 데이터베이스 구축 전에 E/R Diagram을 재정의하면서 데이터로부터 Entity set, relation, attribute등을 정의하였다. Entity set의 attribute 정의 시, primary key와 foreign key를 정의하여 테이블 간 join이 가능하게 했다. 데이터베이스를 구축한 후, 먼저 Entity set간 relationship이 정의한대로 올바르게 작동하는지 확인하기 위해 동일한 서비스 사용자가 여러 개의 컨텐츠 및 리뷰에 접근할 수 있고, 여러 개의 리뷰를 남겼음을 확인하였으므로 기존에 설계한대로 데이터베이스가 잘 작동하는 것을 알 수 있었다.

또한 해당 데이터베이스를 통해 사용자가 특정 영화에 대한 리뷰와 평균평점을 확인할 수 있는 기능이 제공된다.

References

1. Sandvine, “Global Internet Phenomena Report, Spring 2011,” http: //www.sandvine.com/news/global broadband trends.asp, 2011.
2. Liu, Z., & Park, S. (2015). What makes a useful online review? Implication for travel product websites.
3. Tourism management, 47, 140-151.
4. Yoo, K. H., & Gretzel, U. (2008). What motivates consumers to write online travel reviews?. Information Technology & Tourism, 10(4), 283-295.
5. Ye, Q., Law, R., Gu, B., & Chen, W. (2011). The influence of user-generated content on traveler behavior: An empirical investigation on the effects of e-word-of-mouth to hotel online bookings. Computers in Human behavior, 27(2), 634-639.
6. Lucking‐Reiley, D., Bryan, D., Prasad, N., & Reeves, D. (2007). Pennies from eBay: The determinants of price in online auctions. The journal of industrial economics, 55(2), 223-233.
7. Cabral, L., & Hortacsu, A. (2010). The dynamics of seller reputation: Evidence from eBay. The Journal of Industrial Economics, 58(1), 54-78.
8. Chevalier, J. A., & Mayzlin, D. (2006). The effect of word of mouth on sales: Online book reviews. Journal of marketing research, 43(3), 345-354.
9. Boot, P. (2017, August). A Database of Online Book Response and the Nature of the Literary Thriller. In DH.