

Article

[Alex Litkovets](#) · Apr 10, 2017 5m read

## iKnow Review Analyzer (iKRA)



## Introduction

We used the InterSystems iKnow technology to create a review assessment system called iKnow Reviews Analyzer (iKRA). Some information about the prototype of the system can be found [here](#). iKRA analyzes users' text reviews and automatically rates the object being reviewed. This functionality may come in very handy on e-commerce sites, forums or collections of media content – in other words, everywhere where people discuss products, places or services, for example.

## What does the solution do?

iKnow Reviews Analyzer analyzes the domain area, be it the online sales of home appliances or hotel room booking services for travelers. To get the results of this analysis, you need to perform the following steps:

- collect reviews in the designated domain area;
- create dictionaries – databases of words for calculations;
- create an area for loading and analyzing data;
- start model calculations;
- drink coffee / wait;
- look at the results.

# Use case

Here ' s how it looks I reality... Let ' s use smartphone reviews as an example and pick 5 manufacturers:

- Apple;
- Huawei;
- LG;
- Meizu;
- Samsung.

Let ' s assume that we are interested in two models from each of these vendors. Let ' s download 50 reviews for each of the selected models – the total will be 500. Reviews will come from [kimovil](#).

We ' ll save each review to a separate file and use the following file organization scheme (Figure 1):

Имя	Дата изменения	Тип
Apple_iPhone_5s 01 (7.7).txt	01.04.2017 16:31	Файл "TXT"
Apple_iPhone_5s 02 (10).txt	01.04.2017 16:31	Файл "TXT"
Apple_iPhone_5s 03 (9.3).txt	01.04.2017 16:31	Файл "TXT"
Apple_iPhone_5s 04 (8.7).txt	01.04.2017 16:31	Файл "TXT"
Apple_iPhone_5s 05 (9).txt	01.04.2017 16:31	Файл "TXT"
Apple_iPhone_5s 06 (9.9).txt	01.04.2017 16:31	Файл "TXT"
Apple_iPhone_5s 07 (8.9).txt	01.04.2017 16:31	Файл "TXT"
Apple_iPhone_5s 08 (9.8).txt	01.04.2017 16:31	Файл "TXT"
Apple_iPhone_5s 09 (8.8).txt	01.04.2017 16:31	Файл "TXT"
Apple_iPhone_5s 10 (7).txt	01.04.2017 16:31	Файл "TXT"
Apple_iPhone_5s 11 (8.8).txt	01.04.2017 16:31	Файл "TXT"

Figure 1. File location hierarchy

Brackets contain the general smartphone rating specified by the user in the review. It is written to metadata and used afterwards to optimize the calculation algorithm. Source reviews can be found [here](#).

To perform the analysis, you need to create an iKnow domain – a storage of unstructured data. We will not focus on it right now, since this topic is described in detail [here](#).

Once we ' ve created a domain and filled it with reviews, let ' s proceed to analyzing its content. When I choose a smartphone, the following parameters are crucial for me:

- performance;
- quality of communications;
- comfort/ergonomics.

For ease of narration, let me introduce the following notions:

- Category – a rateable parameter;
- Functional (f) marker – a term that characterizes a parameter/category being rated;
- Functional dictionary – an array of f-Markers;
- Emotional (e) marker – a word characterizing the attitude of the reviewer to the object of review;
- Emotional dictionary – an array of e-Markers.

Let ' s use the selected characteristics to create a functional dictionary, where f-Markers (determiners) are assigned

to each of the specified categories. For example, the “ performance ” category is likely to contain markers like “ speed ” , “ processor ” , “ memory ” , “ performance ” , “ core ” and so on. All f-Markers are saved to a special file. Figure 2 shows an example of a “ Performance ” category:

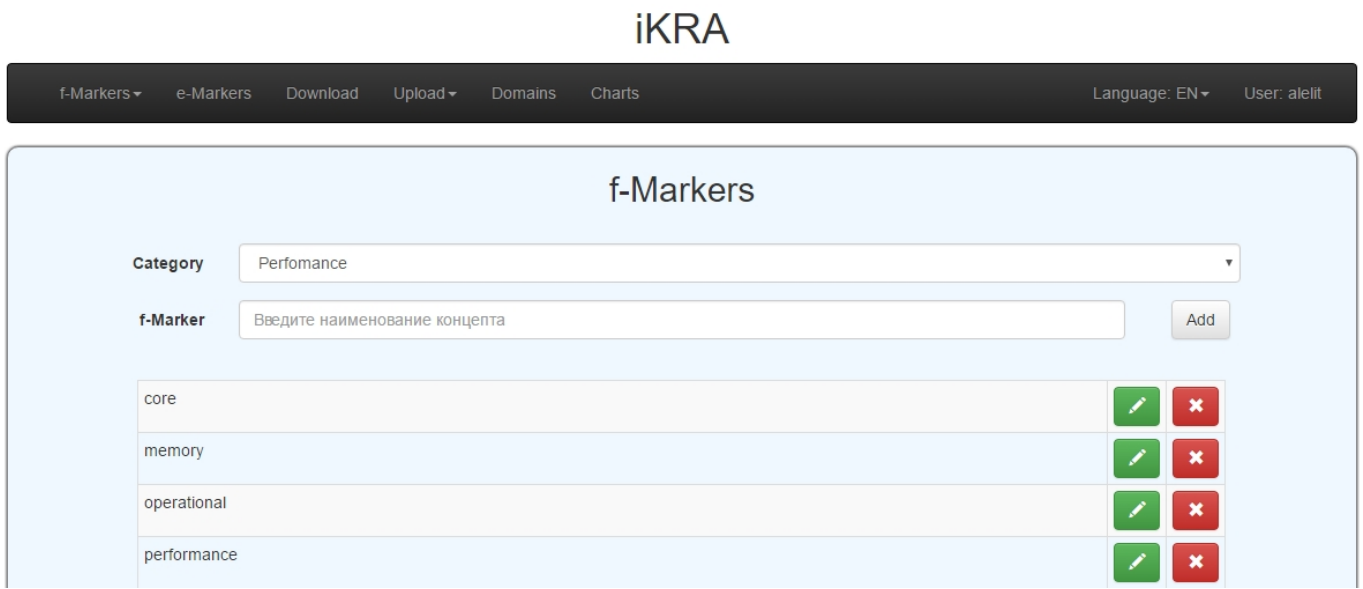


Figure 2. f-Markers

After that, we will create a dictionary of emotions by filling it with corresponding e-Markers. It ’ s impossible to provide a complete list here, but those would be words like “ good ” , “ comfortable ” , “ liked ” , “ issues ” , “ problems ” . Markers define a positive or negative context of every sentence in a review. Each e-Marker will have a numeric value assigned to it. For convenience, let ’ s use +1 for positive markers and -1 for negative ones. All e-Markers are also saved to a special file. Figure 3 shows an example of a set of e-Markers:

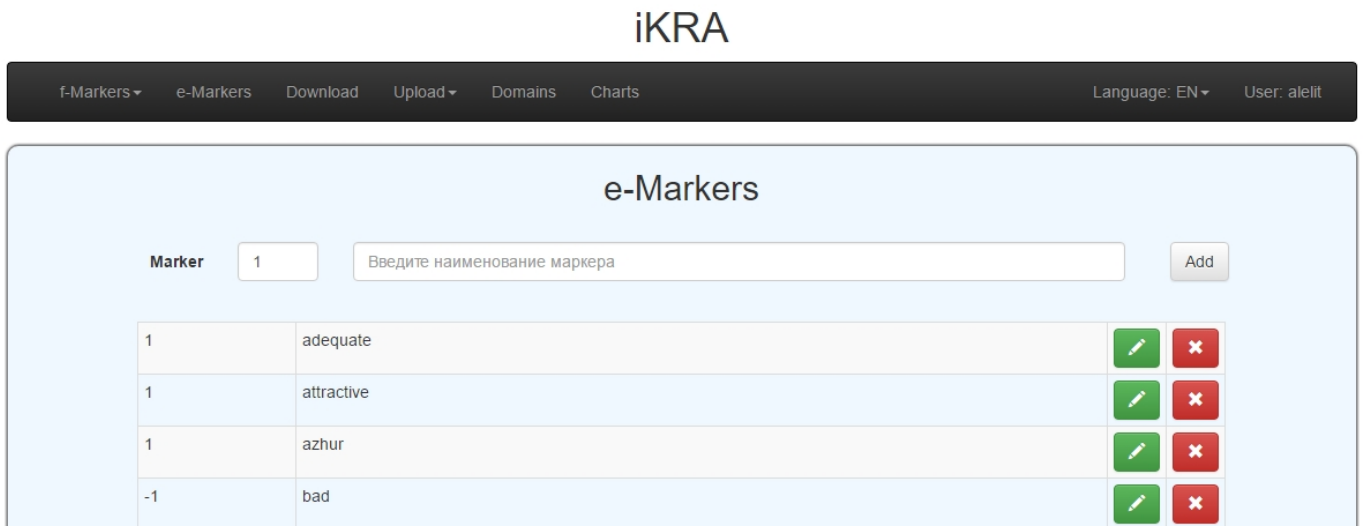


Figure 3. e-Markers

Once the dictionaries are ready, we can calculate the ratings. To do this, select the necessary domain on the “ Domains ” tab and click the “ Calculate ” button (Figure 4):

iKRA

f-Markers ▾e-MarkersDownloadUpload ▾DomainsCharts

Language: EN ▾User: alelit

Domains


Domain


Phones ▾


Calculate


Figure 4. Rating calculation


To view the result, open the `ikra.Dictionary.MarksUnit` class table containing ratings for each smartphone model or use the `ikra.Dictionary.MarksReview` class that contains ratings for each review. Information is shown on the management portal. Let ' s select the SQL section to view the necessary table. Figure 5 demonstrates the viewing of an `ikra.Dictionary.MarksUnit` class table.

 Home

 DeepSee

 System Operation

 System Explorer

 System Administration

Classes

SQL


Routines

Globals

Tools »

iKnow »

SQL



Manage data on this system from a relational point of view.

Go

Add to favorites

System Resource(s)  
%Development

Custom Resource  
-

Assign

CtgID	DomainID	Mark	UID	UName
1	1	9.08	1	Samsung_Galaxy_S6
2	1	10.00	1	Samsung_Galaxy_S6
3	1	9.30	1	Samsung_Galaxy_S6
1	1	10.00	2	Samsung_Galaxy_S7
2	1	10.00	2	Samsung_Galaxy_S7
3	1	10.00	2	Samsung_Galaxy_S7

Figure 5. Viewing the `ikra.Dictionary.MarksUnit` table

Let ' s use DeepSee to check out the result. We ' ve created a cube that uses the results of rating calculation by category and have built a chart for each analyzed smartphone model (Figure 6):

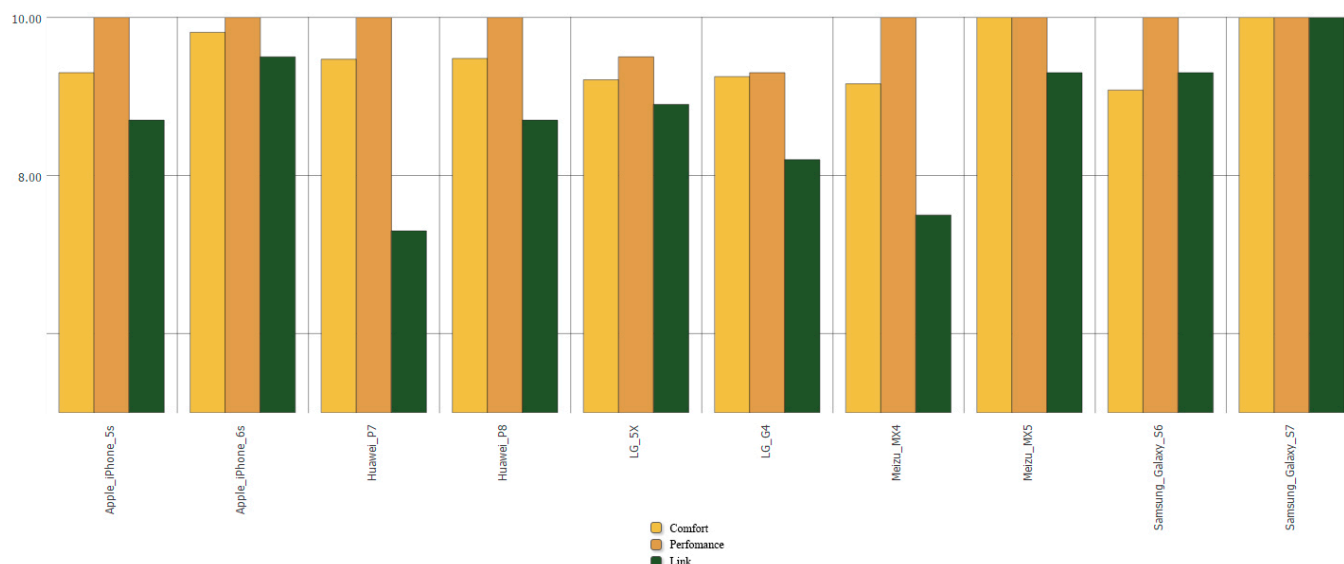


Figure 6. Ratings chart by category

## What if we add another category?

In the past, if you wanted to rate categories, you 'd have to specify the corresponding class property manually. It was inconvenient, since when categories or their count changed during the analysis of new domain areas, you had to make corrections in the code, which wasn 't the most fun and productive use of time. To avoid this, we have considered two solutions:

1. Reservation of a large number of class properties;
2. Use of a database.

The first option allows us to forget about the constantly changing number of categories and not care about the database structure. However, it 's not very convenient to store such a large number of properties, and there are no guarantees that the number of rateable parameters will remain unchanged. Therefore, we decided to go in a different direction.

The second option solves the problem with an undefined number of categories and does not require a fixed amount of memory for storing each class instance. When using a database, the system easily adapts to analyzing any domain area with any number of categories.

The advantages of the second approach convinced us to use it in iKRA.

## Adding a new category

“ And then I realized that I needed to rate another parameter of my smartphone – camera! (If you are into catching Pokemons, do it in style) ”

Adding a new category is easy – all you need to do is to change the content of the functional dictionary and add a new category name (Figure 7).

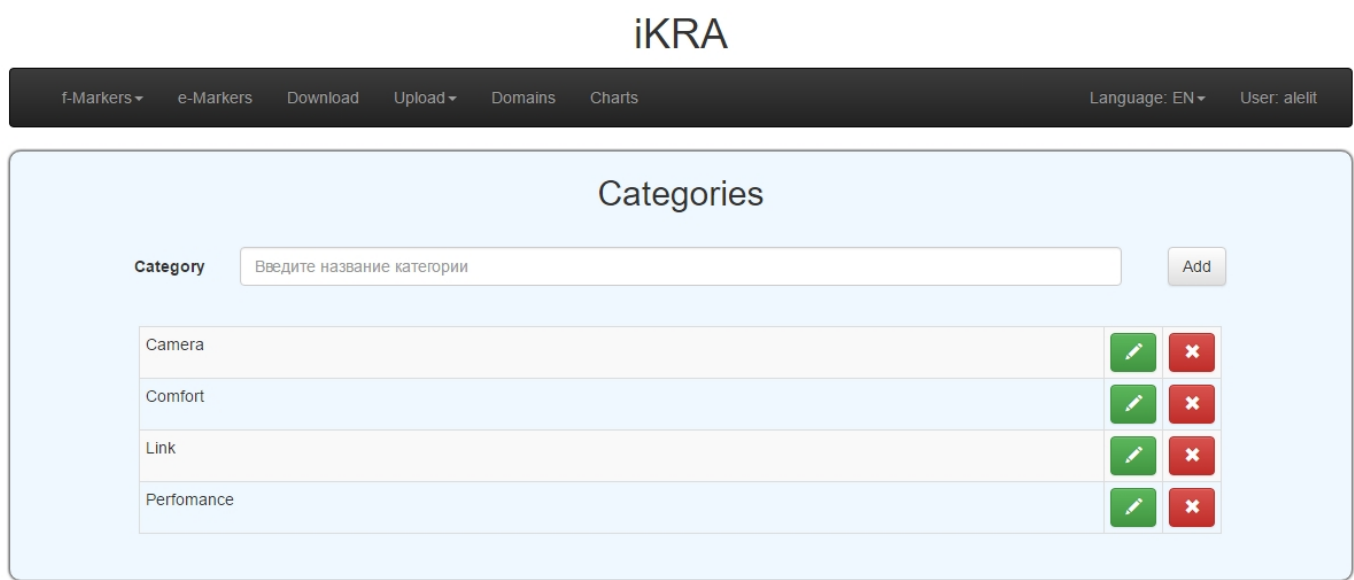


Figure 7. Adding a “ Camera ” category

Let ’ s define a category by adding f-markers on the corresponding tab (Figure 2).

Select the necessary one on the domains tab and start the analysis (Figure 4).

Let ’ s wait for it to finish and then proceed to viewing (Figure 8):

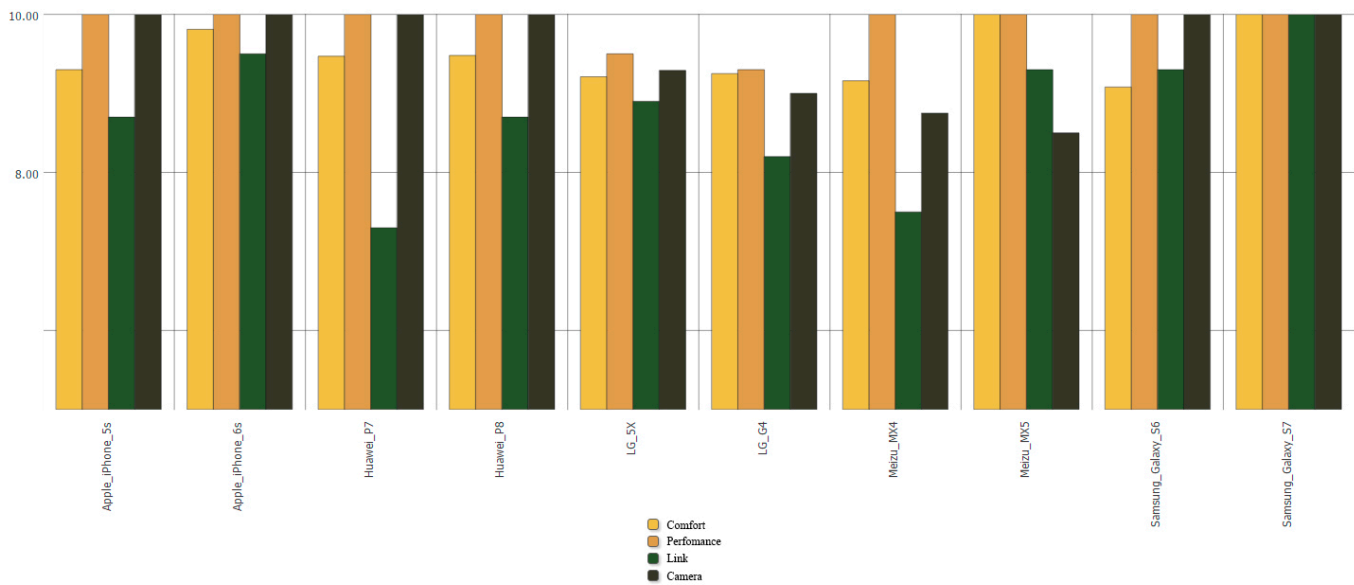


Figure 8. The updated category rating chart

Hooray! We ’ ve easily added a new category and rated it.

## To be continued

We will now be able to rate any product category quickly and without re-writing a line of code. All we need to do is to set up a dictionary and start the analysis. The complex part is the loading of reviews to the database, but we will cover this topic in a separate article.

[GitHub link](#)

[#Databases](#) [#InterSystems](#) [IRIS BI \(DeepSee\)](#) [#Unstructured Data](#) [#InterSystems Natural Language Processing \(NLP, iKnow\)](#)

---

Source URL: <https://community.intersystems.com/post/iknow-review-analyzer-ikra>