PHYSICS TODAY

ANALYSIS OF READER'S FEEDBACK

An analysis of the Nov.2016 Survey Report

Yu (Lucida) Wu, Marketing Intern

Lijun (Reggie) Yang, Co-worker

July 2019

Table of Contents

Executive Summary

In this report, we analyzed customer responses to the open-ended questions in the latest *Physics Today* reader survey (Nov. 2016) with machine learning methods.

We are particularly interested in understanding:

1. What our readers like most about *Physics Today*
2. What our readers like least about *Physics Today*
3. What other sources readers refer to besides *Physics Today*
4. What topics our readers are interested in

We found

1. Current events and new discovery, the diversity of article topics, history of physics and people were the features that our readers valued most
2. Our readers complained most about the following topics: ‘too technical’, ‘no time to read’, ‘too political’, and ‘job ads’
3. The top five resources that our readers referred to, besides *Physics Today*, were Science, Nature, New York Times, Scientific American, and arXiv
4. The popular topics among our readers were: climate change, science policy, recent research results and research funding, education events and support, and international collaboration

Introduction and Background

Our survey data comes from EdStudy Report conducted by Signet Research, Inc. The surveys are conducted every two to three years to measure Readership, Interest, Usefulness and Receptiveness of the editorial content of published issues. Valuable marketing insights can also be generated based on readers feedback.

The questions of the survey are mainly divided into two parts:

1. Multiple-choice questions;
2. Open-ended questions.

Statistical data can be generated easily with the first part. However, analyzing considerable text data could be extremely tricky for traditional statistics. People talking differently might express the same theme, but it is nearly impossible for people to interpret and count the response with

vast amount of data; direct descriptive statistics also seem less useful when handling the unstructured text data. Luckily, with the development of machine learning, algorithms in Natural Language Processing provide much better solutions.

Data and Methodology

Data

By the closing date of January 19, 2017, 1,434 responses were received at a 10.1% response rate. This report has been weighted to reflect the Association composition of the *Physics Today* circulation as follows:

Table 1: Distribution of responses

Society	**% of Total *Physics Today* Circulation**
APS	41%
AAPT	6%
OSA	15%
AAS	5%
AVS	3%
ASA	5%
AAPM	5%
ACA	1%
SOR	1%
SPS	3%
AMS	10%
Direct Order	3%

Methodology

1. Text Preprocessing - transforming text, removing stop words, tokenization, lemmatization
2. TF-IDF - convert the unstructured text data into structured document-term matrix

3. Hierarchical Clustering with Cosine Similarity as the distance matrix was applied to cluster similar readers' reviews (visualization in Appendix Graph 2)
4. Latent Dirichlet Allocation (LDA) was applied to segment and explain reviews by latent topics (Full result in Appendix Table 1 and Table 2)

Key Findings

The statistics of the readers' reviews in the survey of 2016 are as follows:

Table 2: Data Scale

Question (Dataset)	Response	Valid Response
What do you like most about PT?	392 reviews	381 reviews
What do you like least about PT?	307 reviews	287 reviews
Other sources besides PT	387 reviews	387 reviews
Topics readers interested in	172 reviews	172 reviews

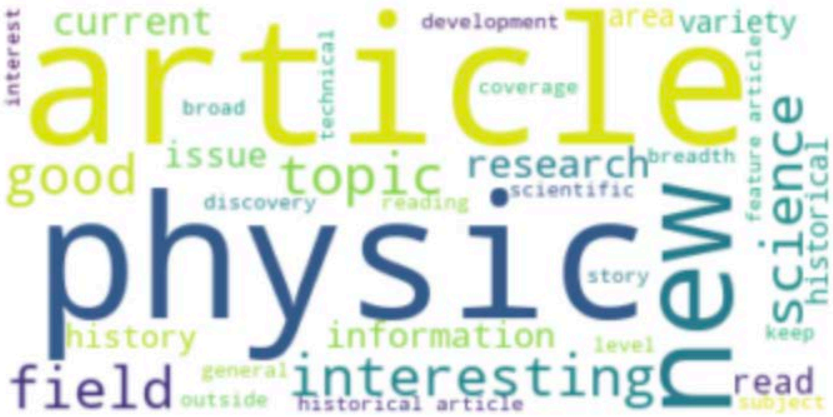
Based on the four datasets above, we extracted and analyzed the common themes people talked about to get deeper understanding of our readers.

Question 1. What do you like most about *Physics Today*?

Analyze on Word Level: WordCloud

Firstly, we utilized WordCloud to extract and visualize the words recipients commonly mentioned about. The larger the words, the higher frequency the words were mentioned.

Graph 1: Original WordCloud for Question 1



The original WordCloud reflected all commonly used words. As we expected, people mostly talked about physics-related topics and various articles in *Physics Today*. However, people's preference was not reflected clearly due to many frequent but less meaningful words like 'good', 'interesting', 'field'. To extract more specific features that our readers valued most, words with less meaning were removed, and the cleaned WordCloud was generated.

Graph 2: Cleaned WordCloud for Question 1

The new WordCloud was much clearer.

1. "historical" and "history" revealed readers' high preference and their endorsement of the quality of physics history articles. Examples in the reviews could be "history of physics", "historical reviews of topics/people";
2. "new", "current" and "discovery" revealed readers' great interest in the "Search and Discovery" part, which keeps them informed about current events and new discoveries in

physics. Examples could be “cutting edge news”, “reporting on new discoveries and work of real or potential importance”;

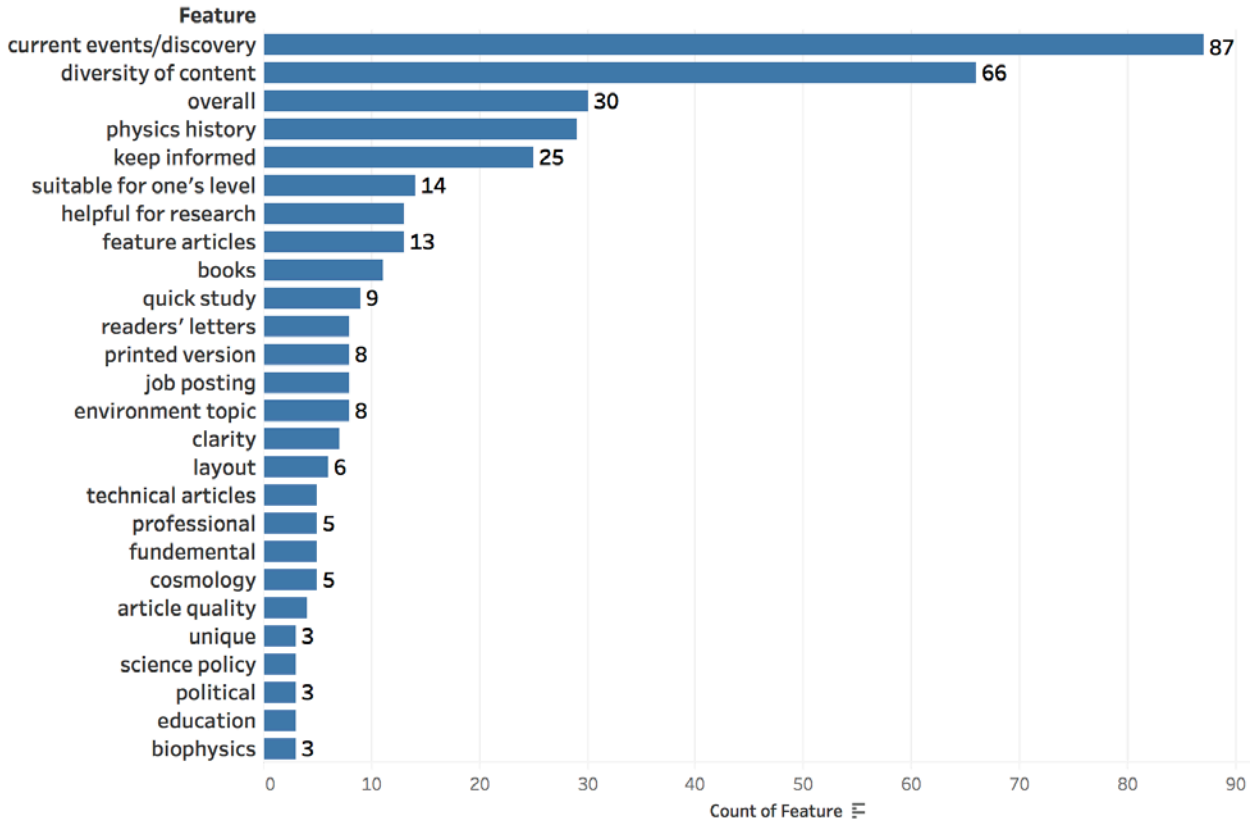
3. “variety”, “information”, “outside”, “broad”, “breadth” reviewed that readers value the diversity and breadth of the topics in *Physics Today*, which broad their horizons outside their own fields. Examples could be “wide variety of branches of physics”, “I enjoy the breadth of topics considered”;
4. “feature (article)”, “search (and) discovery”, “job” and “book” reveal readers’ favorite categories of *Physics Today*.

Clustering Whole Sentences: Hierarchical Clustering

Besides word-level analysis, with the results of hierarchical clustering, we managed to put the reviews with similar meanings into the same clusters and count the scale of each cluster to find out the most popular features of *Physics Today*.

With a few manual modifications on the original clustering results, we finally got 35 clusters of similar reviews. Clusters larger than three are shown in the following bar chart (Graph 3). The full-size bar chart (Appendix Graph 1) and the visualization of hierarchical results (Appendix Graph 2) are in the Appendix.

Graph 3: Popular features of Physics



According to this bar chat, the current events/news/discoveries of physics, most of which came from the "search and discovery" part, were the most attracting contents in *Physics Today*. 87 (22.83%) responses mentioned these contents as their favorite features of the magazine. The diversity of topics and subjects is another attracting feature. 66 (17.32%) respondents showed their appreciation in the diversity and breadth of articles. Many people also mentioned the high quality and attractiveness of the historical stories and people in physics.

Overall speaking, most people liked *Physics Today* because it kept them informed and updated to the current physics.

1. For each part of *Physics Today*, "Search and Discovery", "Feature Articles", "Books", "Quick study", "Readers' Forum", "Job opportunities" were the relatively popular categories. (rank above has meaning)
2. Quite a few responses said *Physics Today* were helpful for their researches, and they liked the property of "technical". However, lots of people also complained that the articles were too technical for them (Shown later in the "What do you like least about *Physics Today*?" part).

3. Printed version was regarded as one obvious advantage for some subscribers.
4. Overall speaking, the results of this algorithm corresponded well to the results of WorldClouds.

Topic Modeling: Latent Dirichlet Allocation (LDA)

Last, with LDA, we succeeded to explain most reviews with a small number of topics and check the stability of the former algorithm (hierarchical clustering).

We generated 7 topics with LDA to segment different responses about what they liked most about *Physics Today*. The details of each topic are shown in Appendix Table 1.

Table 3: Topics and Counts for Question 1

Topic	Theme	Count
1	information variety; helpful for researches	36
2	breadth of topics; new discovery;	42
3	resent news and development in physics	69
4	recent news; article diversity; article quality	27
5	physics history; feature articles;	52
6	book reviews; physics history; job posting	48
7	recent news and development; physics history	34
8	others	84

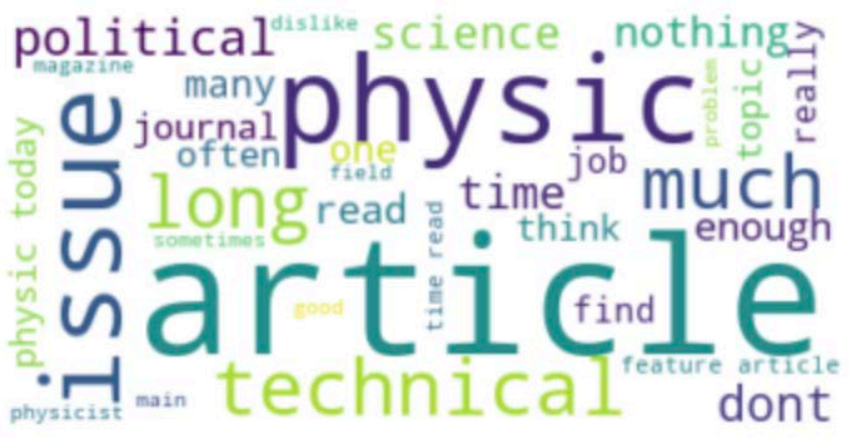
There might be a few overlaps among these topics. Overall speaking, the results of the topic modeling corresponded to the results of hierarchical clustering (the first algorithm) and WordCloud. With the three angles above, our finial conclusions could be more trustworthy.

Question 2. What do you like least about *Physics Today*?

Analyze on Word Level: WordCloud

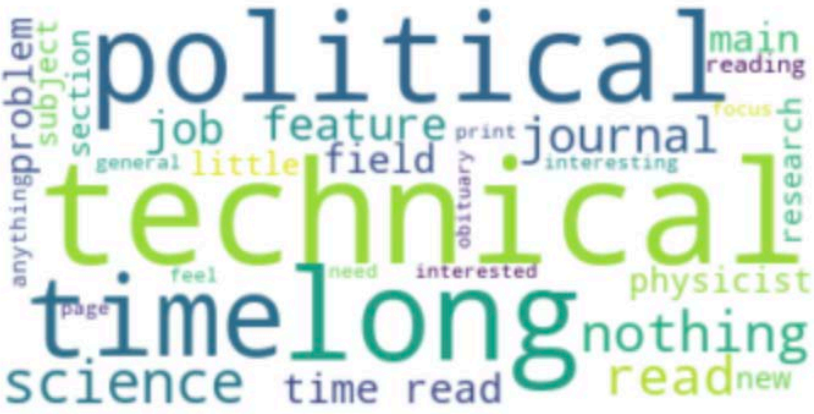
Firstly, we utilized the WordCloud to extract and visualize the words that recipients commonly mentioned about.

Graph 4: Original WordCloud for Question 2



After moving out some less meaningful but frequent words, the result became much clearer.

Graph 5: Cleaned WordCloud for Question 2



When talking about dislike features of *Physics Today*, complains like “articles are too technical”, “too much emphasis on political issues”, “some articles are too long” and “no time to read it” were most frequently mentioned.

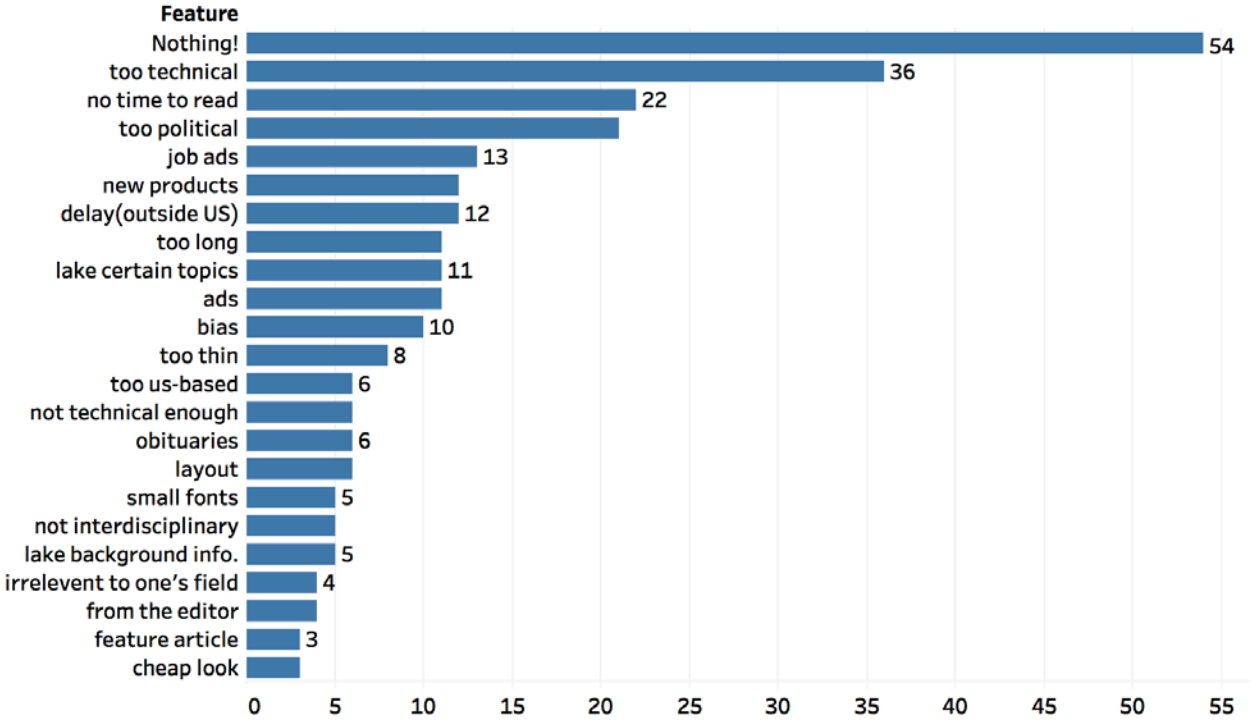
However, the relatively big word “nothing” reviewed that quite a few readers were actually very satisfied with the current *Physics Today*. In addition, the responses amount of question 2 (least like) was less than it of question 1 (most like), this phenomenon also reviewed, to certain degree, that our readers did not have much to say about how they dislike the magazine.

Clustering Whole Sentences: Hierarchical Clustering

With the results of hierarchical clustering, we managed to put most reviews of similar meanings into the same clusters, and count the scale of each cluster to find what our readers like least about *Physics Today*.

We finally got 38 clusters of similar reviews. Clusters larger than three were shown in the following bar chart. The full-size bar chart (Appendix Graph 3) and the visualization of hierarchical results (Appendix Graph 4) are in the Appendix.

Graph 6: Complained Features of *Physics Today*



54 (18.82%) responses actually had nothing to complain about current *Physics Today*. Example comments could be “Nothing.”, “I love it all”, “Generally terrific.”

Except these reviews, the most common features that people are not satisfied about are:

1. Articles are too technical (though a few responses also said the articles were not technical enough);
2. The magazine contained too much information, or some articles were too long. They did not have enough time to read each issue (though a few responses also said the magazines were too thin).
3. Too much emphasis was put on political issues.
4. Job opportunities
5. New products
6. The phenomenon of delay (for printed version and readers mostly outside US).
7. Lake certain specific topics that the readers were interested, like AMO, lasers, representation of women
8. Advertisings
9. Bias in certain articles, especially relative to climate change. One example could be “Overwhelming bias for anything/anyone speaking favorably about issues of global climate change.”

Topic Modeling: Latent Dirichlet Allocation (LDA)

Last, with LDA, we succeeded to explain most reviews with a small number of topics, and check the stability of the former algorithm (hierarchical clustering).

We generated 10 topics with LDA to segment different responses about what they liked least about *Physics Today*. The details of each topic are shown in Appendix Table 2.

Table 4: Topics and Counts for Question 2

Topic	Theme	Count

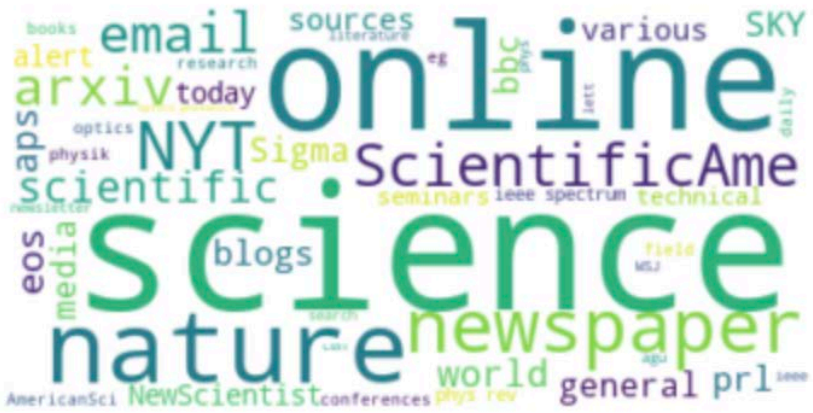
1	too political; bias	21
2	too technical; articles too long	23
3	lake certain topics; advertising	23
4	irrelavent to one's field; too political	28
5	new product sectioon; too political	22
6	overall good	16
7	some articles too short; too peripheral(not technical enough)	13
8	job posting; obituary	19
9	overall good; fonts too small	18
10	repeating old arguments; equipment ads	26
11	others	98

Overall speaking, the results of the topic modeling corresponded with the results of hierarchical clustering (the first algorithm) and the results of WordCloud. With the three angles above, our finial conclusions could be more trustworthy.

Question 3. From what other sources do you learn about work that is later reported on "Search & Discovery"?

After a series of preprocessing, we generated a WordCloud, which could reveal the frequency of the source names (e.g., Science, Nature, NYT, Scientific American) and the frequency of the channels that our readers usually got information from (e.g., online, newspaper, email).

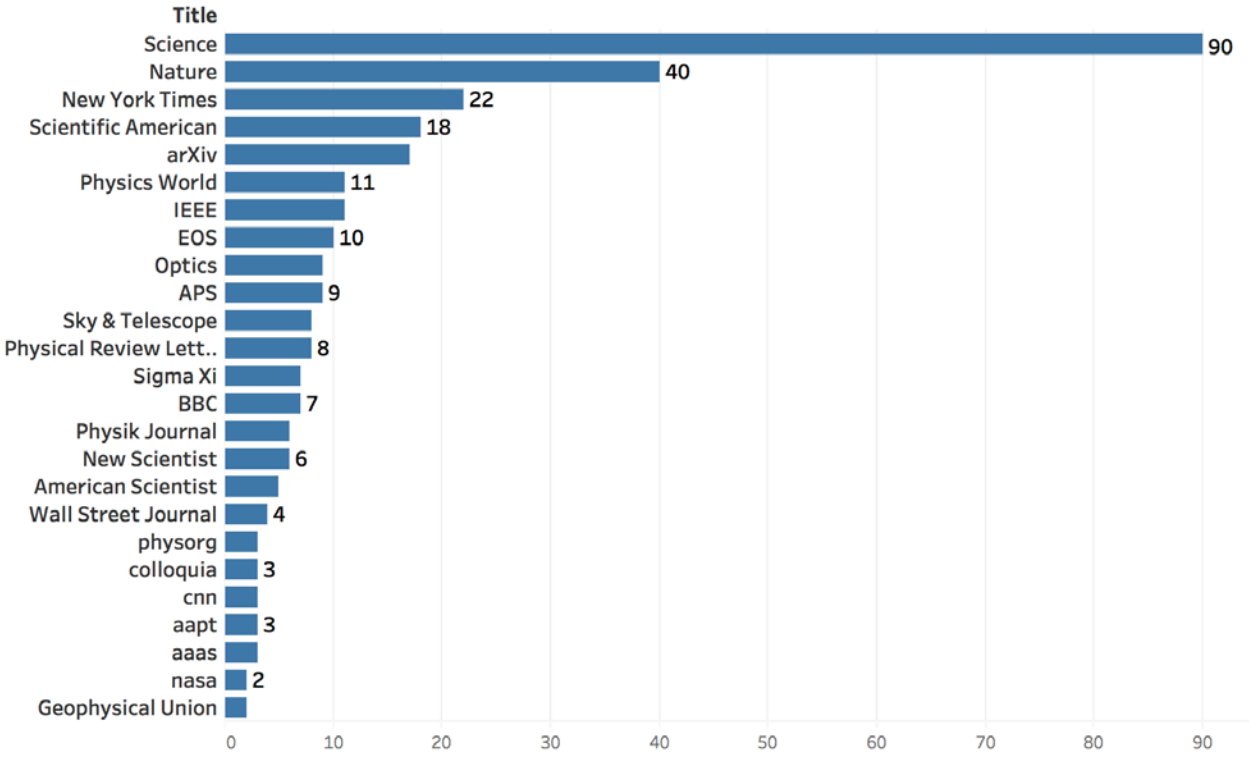
Graph 7: WordCloud of Popular Source



Then we identified

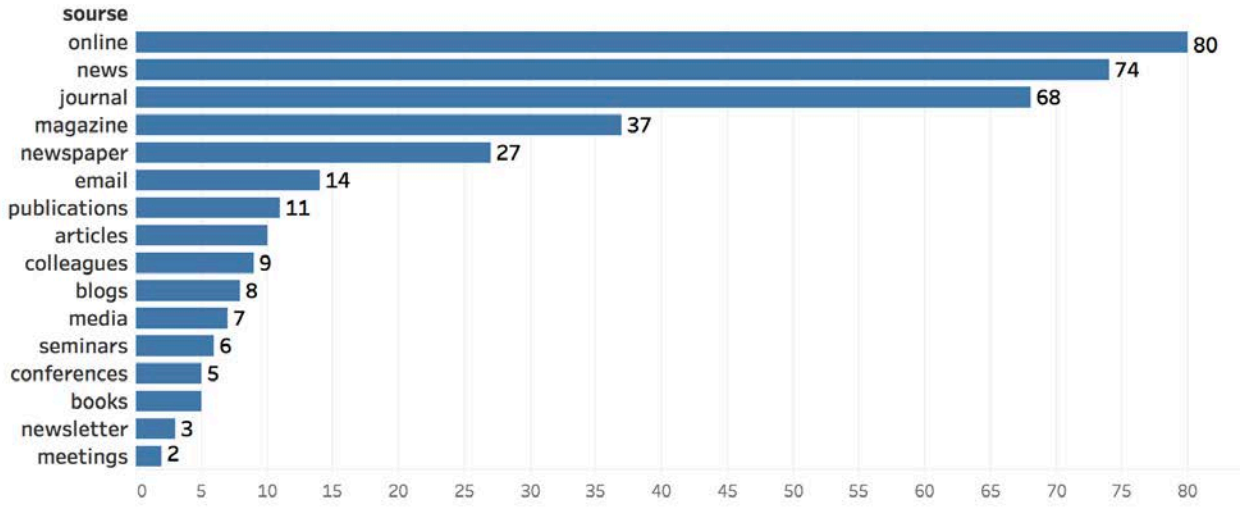
1. the most popular magazines/journals/social media that our readers referred to, besides Physics Today:

Graph 8: Other Sources Besides *Physics Today*



2. the most popular channels that our readers gain information:

Graph 9: Channels



All words whose frequency are larger than three in the whole dataset are shown in Appendix Graph 5.

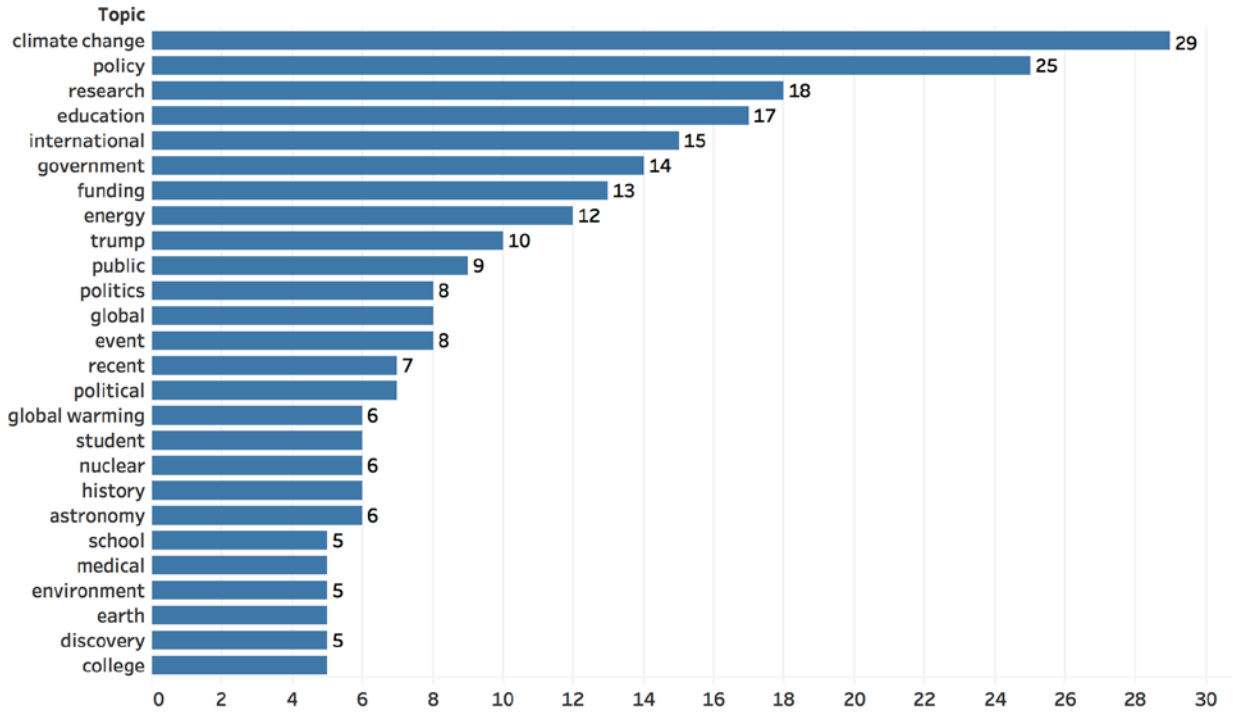
Question 4. What topics or issues should be covered more often in Issues & Events?

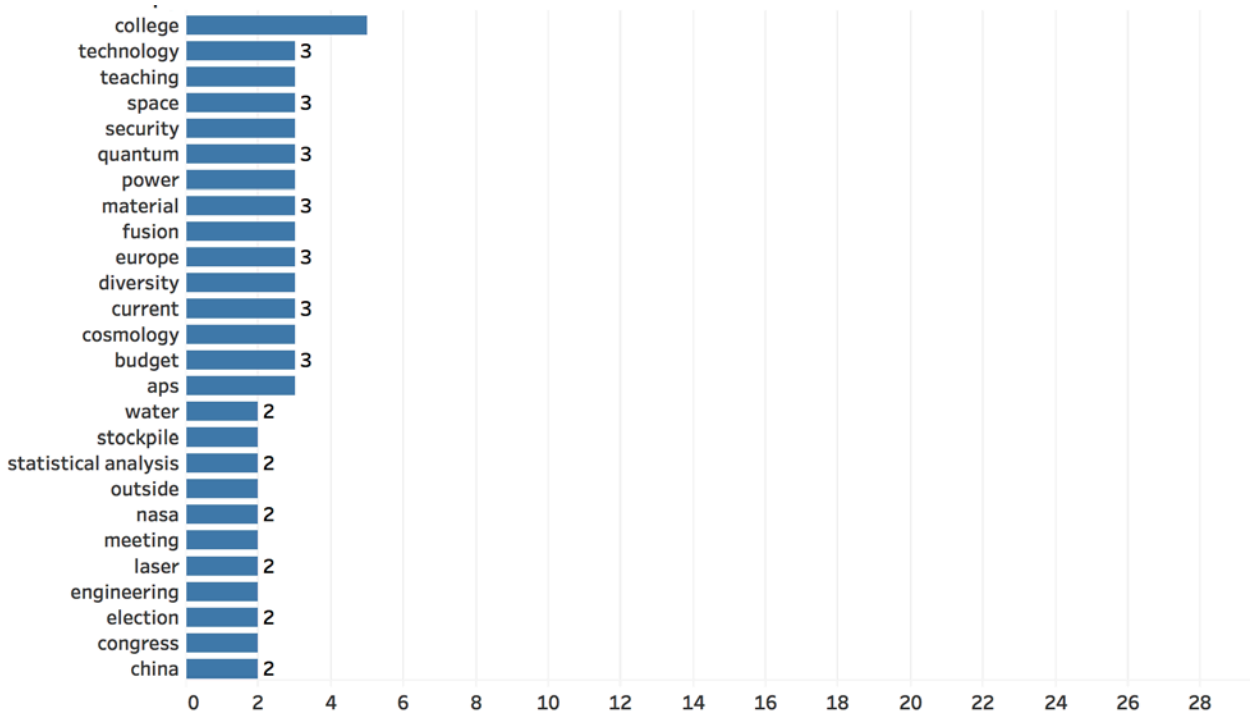
After a series of preprocessing, we generated a WordCloud, which could reveal the frequency of the popular topics among the responses.

Graph 10: WordCloud of Popolar Topics

Then we identified the most popular topics among our readers and displayed it clearly with a bar chart.

Graph 11: Popular Topics





Conclusions and Recommendations

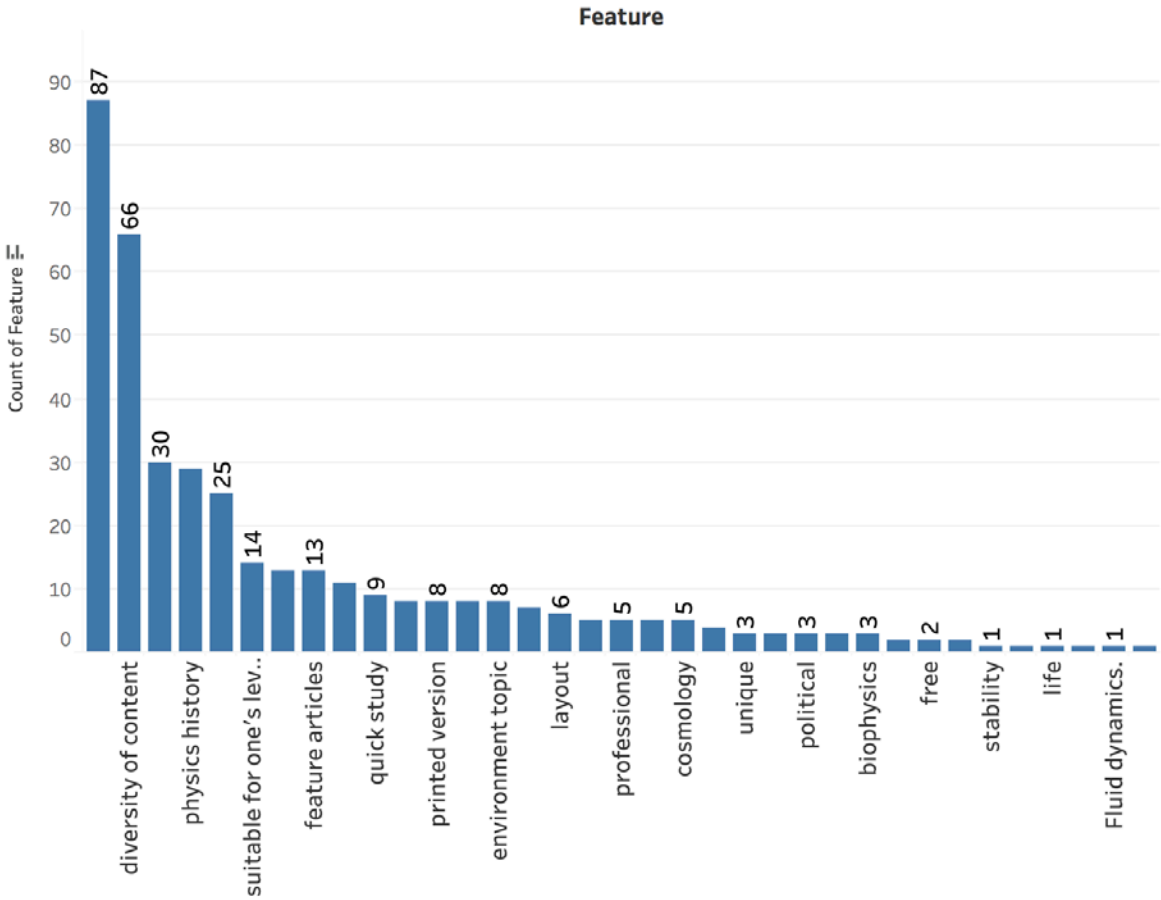
1. People liked *Physics Today* mostly because it helped them stay updated in the field of physics by providing the cutting-edge news, discoveries, events and overviews of researches. People considered it as a professional source of information. When promoting the magazine or creating new contents in AIP/*Physics Today* social media, we could refer to the features that our readers like most.
2. Articles in physics history seemed to have done a really good job.
3. *Physics Today* actually had quite a few readers who were not professional physics learners. They might be high-school physics teachers, or just had interest in physics. Many articles in *Physics Today* were too technical for them.
4. Modifications might be done based on the features people like least. But it's really hard to satisfy everyone. For example, a lot of people said they did not have time to read the monthly issues, but some people complained that the magazines were too thin.

Limitations

1. If we have more sufficient data, the unsupervised learning algorithms applied in this analytics could work better, and some more advanced supervised learning algorithms could also be applied.
2. The survey data is lagged. Some information might only be valid for that time (2016)

Appendices: Tables, Exhibits, Figures

Graph 1



Graph 2

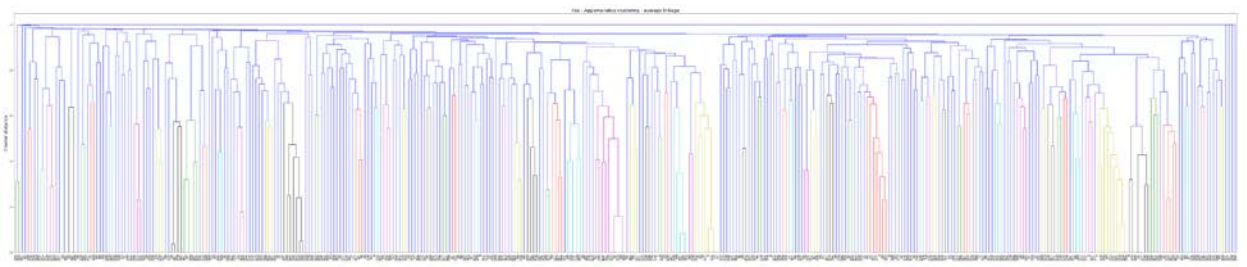
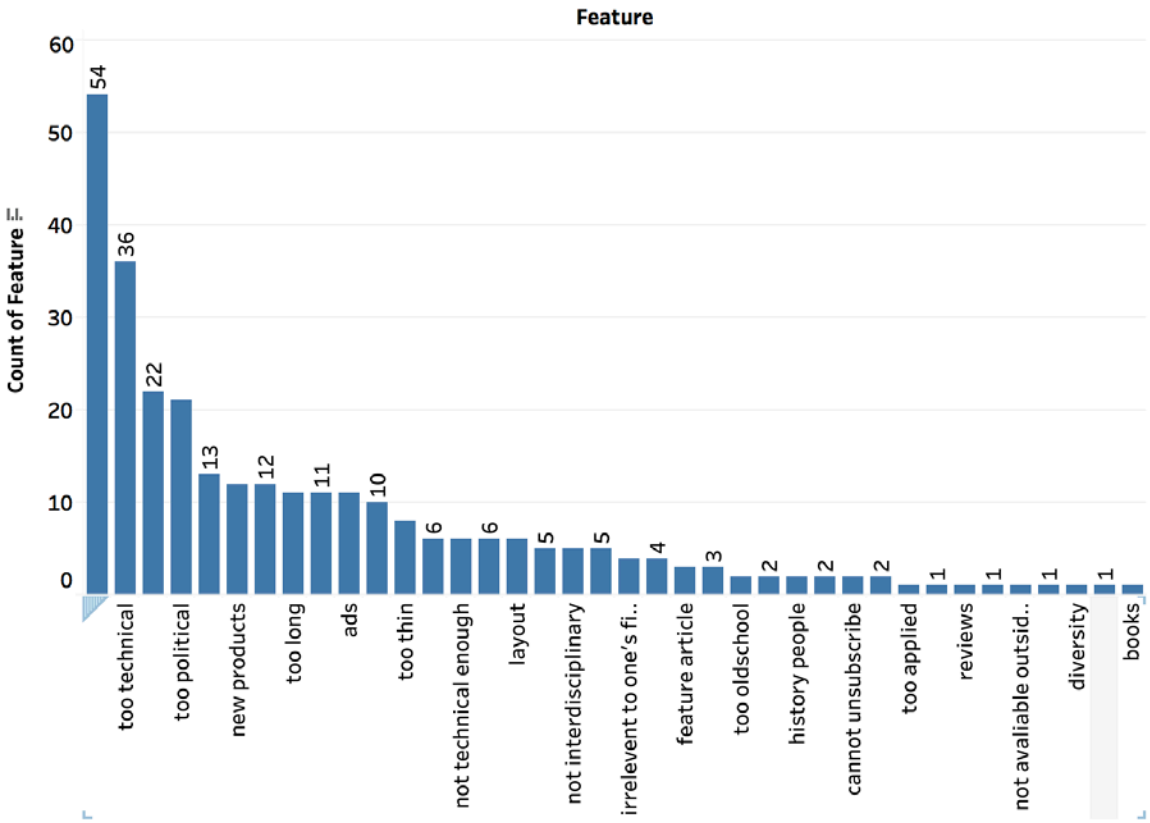


Table 1

Topic	Detai
1	0.026*"variety" + 0.023*"information" + 0.017*"research" + 0.015*"physic" + 0.012*"outside" + 0.012*"issue" + 0.010*"informative" + 0.010*"timely" + 0.010*"physicist" + 0.010*"read"
2	0.036*"physic" + 0.031*"broad" + 0.024*"range" + 0.023*"covered" + 0.022*"discovery" + 0.014*"scientific" + 0.014*"historical" + 0.013*"search" + 0.012*"like" + 0.012*"breadth"
3	0.070*"physic" + 0.029*"development" + 0.029*"new" + 0.018*"field" + 0.018*"coverage" + 0.017*"science" + 0.015*"current" + 0.015*"keep" + 0.013*"research" + 0.013*"technical"
4	0.045*"science" + 0.029*"physic" + 0.023*"news" + 0.021*"recent" + 0.016*"field" + 0.015*"physicist" + 0.015*"variety" + 0.011*"research" + 0.010*"new" + 0.009*"quality"
5	0.025*"good" + 0.020*"historical" + 0.020*"feature" + 0.017*"like" + 0.017*"new" + 0.016*"physic" + 0.016*"subject" + 0.014*"something" + 0.012*"letter" + 0.011*"interest"
6	0.034*"book" + 0.031*"review" + 0.028*"physic" + 0.023*"history" + 0.022*"news" + 0.015*"field" + 0.014*"historical" + 0.013*"new" + 0.013*"job" + 0.013*"like"
7	0.029*"physic" + 0.019*"keep" + 0.017*"current" + 0.014*"science" + 0.012*"historical" + 0.012*"date" + 0.012*"piece" + 0.012*"give" + 0.012*"field" + 0.011*"going"

Graph 3



Graph 4

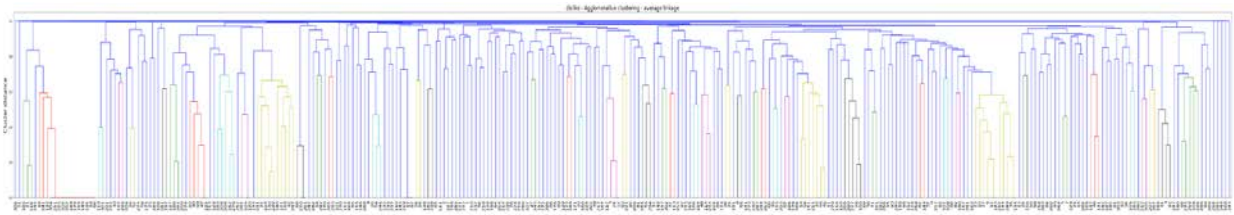
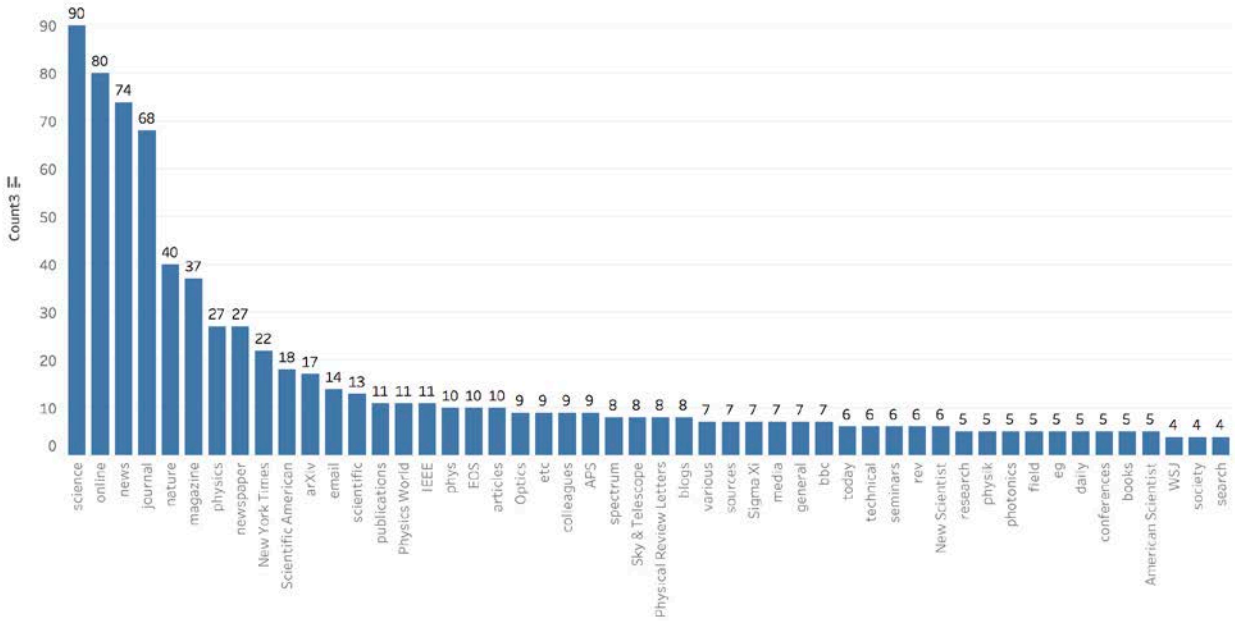


Table 2

Topic	Detail
1	0.062*"political" + 0.028*"much" + 0.017*"fact" + 0.017*"general" + 0.017*"bias" + 0.017*"view" + 0.017*"occasional" + 0.016*"topic" + 0.013*"time" + 0.012*"knowledge"

2	0.136*"time" + 0.040*"technical" + 0.035*"main" + 0.030*"journal" + 0.030*"enough" + 0.029*"long" + 0.017*"many" + 0.015*"topic" + 0.010*"today" + 0.010*"author"
3	0.025*"today" + 0.025*"fine" + 0.019*"dislike" + 0.019*"lack" + 0.019*"review" + 0.019*"ad" + 0.019*"product" + 0.018*"advertisement" + 0.013*"background" + 0.013*"high"
4	0.027*"field" + 0.019*"research" + 0.016*"lot" + 0.016*"change" + 0.016*"political" + 0.016*"topic" + 0.016*"subject" + 0.016*"particular" + 0.015*"climate" + 0.015*"many"
5	0.053*"product" + 0.052*"new" + 0.027*"political" + 0.026*"section" + 0.022*"relevant" + 0.017*"science" + 0.016*"field" + 0.016*"thing" + 0.011*"technical" + 0.011*"interested"
6	0.027*"science" + 0.023*"anything" + 0.023*"journal" + 0.019*"magazine" + 0.019*"feature" + 0.019*"problem" + 0.014*"work" + 0.014*"woman" + 0.014*"year" + 0.014*"photo"
7	0.101*"technical" + 0.035*"short" + 0.021*"format" + 0.014*"line" + 0.014*"peripheral" + 0.014*"prefer" + 0.014*"topic" + 0.014*"theoretical" + 0.014*"reading" + 0.014*"ad"
8	0.039*"job" + 0.037*"much" + 0.030*"opportunity" + 0.024*"science" + 0.020*"obituary" + 0.020*"focus" + 0.011*"area" + 0.010*"summary" + 0.010*"hard" + 0.010*"american"
9	0.084*"nothing" + 0.019*"job" + 0.019*"small" + 0.019*"paper" + 0.014*"print" + 0.014*"interested" + 0.014*"copy" + 0.014*"editor" + 0.010*"section" + 0.010*"enough"
10	0.018*"old" + 0.018*"physicist" + 0.018*"much" + 0.014*"way" + 0.014*"equipment" + 0.009*"climate" + 0.009*"late" + 0.009*"audience" + 0.009*"thinking" + 0.009*"decade"

Graph 5


Count3 F
90
80
74
68
40
37
27 27
22
18 17
14 13
11 11 11
10 10 10
9 9 9 9
8 8 8 8
7 7 7 7 7 7
6 6 6 6 6
5 5 5 5 5 5 5 5
4 4 4
0
10
20
30
40
50
60
70
80
90
science
online
news
journal
nature
magazine
physics
newspaper
New York Times
Scientific American
arXiv
email
scientific
publications
Physics World
IEEE
phys
EOS
articles
Optics
etc
colleagues
APS
spectrum
Sky & Telescope
Physical Review Letters
blogs
various
sources
Sigma Xi
media
general
bbc
today
technical
seminars
rev
New Scientist
research
physik
photonics
field
eg
daily
conferences
books
American Scientist
WSJ
society
search