

INNOPOLIS
UNIVERSITY

● ОПОРНЫЙ
ОБРАЗОВАТЕЛЬНЫЙ
ЦЕНТР



Казанский Государственный
Медицинский Университет

История инфометрии и библиометрии

Хакимов Нияз

niyaz.hakimov@kazangmu.ru



План лекции

История инфометрии и библиометрии



Казанский Государственный
Медицинский Университет

INNOPOLIS
UNIVERSITY

ОПОРНЫЙ
ОБРАЗОВАТЕЛЬНЫЙ
ЦЕНТР

- Инфометрия
- Библиометрия



История инфометрии и библиометрии



Казанский Государственный
Медицинский Университет

INNOPOLIS
UNIVERSITY

ОПОРНЫЙ
ОБРАЗОВАТЕЛЬНЫЙ
ЦЕНТР

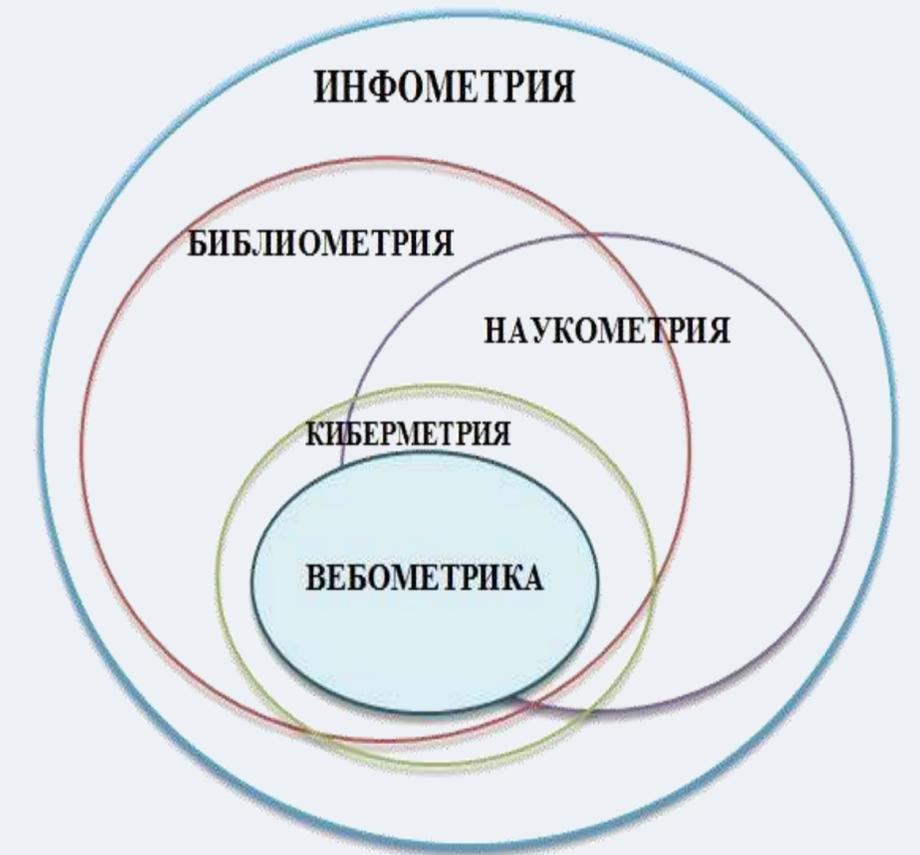
Инфометрия

Инфометрия — научная дисциплина, предметом которой являются измерения количественных характеристик информации.

Термин инфометрия (фр. Infométrie) был предложен в 1979 году французским исследователем О. Накке.

Инфометрия охватывает производство, передачу и использование всех форм информации, независимо от её происхождения

Предметом изучения инфометрии являются измерения количественных характеристик информации



Составляющие части инфометрии и их взаимоотношение

Nacke O. Informetrie: eine neuer Name für eine neue Disziplin. Nachrichten für Dokumentation. 1979. № 30 (6): 219–226.

<https://commons.wikimedia.org/wiki/File:Informetrics.jpg?uselang=ru>



История инфометрии и библиометрии



Казанский Государственный
Медицинский Университет

INNOPOLIS
UNIVERSITY

ОПОРНЫЙ
ОБРАЗОВАТЕЛЬНЫЙ
ЦЕНТР

Части инфометрии

Библиометрия — научная дисциплина, занимающаяся изучением документов на основе количественного анализа первичных и вторичных источников информации с помощью формализованных методов с целью получения данных об эффективности, динамике, структуре и закономерностях их развития.

Вебометрика — научная дисциплина, в рамках которой исследуются количественные аспекты конструирования и использования информационных ресурсов, структур и технологий применительно к Всемирной паутине.

Киберметрия — научная дисциплина, занимающаяся количественным анализом новых возможностей по обработке информации, хранимой в электронном виде.

Наукометрия — научная дисциплина, занимающаяся изучением количественных методов развития науки как информационного процесса.



История инфометрии и библиометрии



Казанский Государственный
Медицинский Университет

INNOPOLIS
UNIVERSITY

ОПОРНЫЙ
ОБРАЗОВАТЕЛЬНЫЙ
ЦЕНТР

Определение информации

- знания о предметах, фактах, идеях и т. д., которыми могут обмениваться люди в рамках конкретного контекста¹
- знания относительно фактов, событий, вещей, идей и понятий, которые в определённом контексте имеют конкретный смысл²
- сведения, воспринимаемые человеком и (или) специальными устройствами как отражение фактов материального или духовного мира в процессе коммуникации³
- запомненный выбор одного варианта из нескольких возможных и равноправных⁴
- сведения о системе, о её структуре и функции, выраженные моделью», а «переработка информации заключается в превращении одних моделей в другие⁵

1- ISO/IEC 10746-2:1996

2 - ISO/IEC 2382:2015

3 - ГОСТ 7.0-99

4 - Кастлер Г. Возникновение биологической организации. — М.: Мир, 1967.

5 - Амосов Н. М. Моделирование сложных систем. — Киев: Наукова думка, 1968. — 88 с.



История инфометрии и библиометрии



Виды информации



История инфометрии и библиометрии



Казанский Государственный
Медицинский Университет

INNOPOLIS
UNIVERSITY

ОПОРНЫЙ
ОБРАЗОВАТЕЛЬНЫЙ
ЦЕНТР

Измерение информации

Ральф Хартли в 1928 г. ввёл понятие «информации» (энтропии) как случайной переменной и предложил измерять информацию логарифмическими мерами: $H = K \log_2[(M)]$

Бит — единица измерения количества информации (0 или 1), Байт = 8 бит, $2^8 = 256$ значений

Трит — логарифмическая единица измерения

1 трит = $\log_2 3 \approx 1,58496$ бит

Трайт = 6 трит

Нат — единица измерения информации, выраженная через натуральный логарифм

Нат равен $\log_2 e \approx 1,443$ бит

Децит — единица измерения информации, выраженная через десятичный логарифм

Децит = $\log_2 10 \approx 3,32192$ бит



Ralph Vinton Lyon
Hartley —
американский
учёный-
электронщик

История инфометрии и библиометрии



Измерения в байтах

ГОСТ 8.417—2002			Приставки СИ			Приставки МЭК		
Название	Обозначение	Степень	Название	Степень	Название	Обозначение		Степень
байт	Б	10^0	—	10^0	байт	В	Б	2^0
килобайт	Кбайт	10^3	кило-	10^3	кибибайт	KiB	КиБ	2^{10}
мегабайт	Мбайт	10^6	мега-	10^6	мебибайт	MiB	МиБ	2^{20}
гигабайт	Гбайт	10^9	гига-	10^9	гибибайт	GiB	ГиБ	2^{30}
терабайт	Тбайт	10^{12}	тера-	10^{12}	тебибайт	TiB	ТиБ	2^{40}
петабайт	Пбайт	10^{15}	пета-	10^{15}	пебибайт	PiB	ПиБ	2^{50}
эксабайт	Эбайт	10^{18}	экса-	10^{18}	эксбибайт	EiB	ЭиБ	2^{60}
зеттабайт	Збайт	10^{21}	зетта-	10^{21}	зебибайт	ZiB	ЗиБ	2^{70}
йоттабайт	Ибайт	10^{24}	йотта-	10^{24}	йобибайт	YiB	ЙиБ	2^{80}



История инфометрии и библиометрии



Казанский Государственный
Медицинский Университет

INNOPOLIS
UNIVERSITY

ОПОРНЫЙ
ОБРАЗОВАТЕЛЬНЫЙ
ЦЕНТР

Инфометрия

Наукометрия – измерение и анализ научной литературы
раздел информатики

- измерение влияния исследовательских работ и академических журналов, научных цитат
- использование измерений в контексте политики и управления
- создала систему порочных стимулов «публикуй или исчезни»



История инфометрии и библиометрии



Казанский Государственный
Медицинский Университет

INNOPOLIS
UNIVERSITY

ОПОРНЫЙ
ОБРАЗОВАТЕЛЬНЫЙ
ЦЕНТР

Законы:

- закон Лотки
- закон Ципфа
- закон Брэдфорда

Работы:

- Дерек Дж. де Солл Прайс
- Джерард Салтон
- Лео Эгге
- Рональд Руссо
- Тибор Браун
- Олле Перссон
- Питер Ингверсен
- Манфред Бониц
- Юджин Гарфилд
- Роберт К. Мертон



История инфометрии и библиометрии



Казанский Государственный
Медицинский Университет

INNOPOLIS
UNIVERSITY

ОПОРНЫЙ
ОБРАЗОВАТЕЛЬНЫЙ
ЦЕНТР

Закон Лотки

- $n_x = n_1 / x^2$
- Закон обратных квадратов, где количество авторов n_x , опубликовавших x статей, является отношением количества авторов n_1 , опубликовавших одну статью, к числу статей x , опубликованных авторами n_x .
- По мере увеличения количества публикуемых статей количество авторов, выпускающих такое количество публикаций, снижается. Количество авторов, опубликовавших две статьи в течение определенного периода времени, составляет $1/4$, публикующих три статьи – $1/9$, публикующих четыре статьи – $1/16$ от количества авторов, опубликовавших одну публикацию и т. д.



История инфометрии и библиометрии



Казанский Государственный
Медицинский Университет

INNOPOLIS
UNIVERSITY

ОПОРНЫЙ
ОБРАЗОВАТЕЛЬНЫЙ
ЦЕНТР

Закон Ципфа

- Закон Ципфа: в физических и социальных науках, ранг-частотное распределение имеет обратную зависимость.
- $P_n = 1/n^a$, где P_n — частота слова, занимающего n -е место, a — показатель степени, a приблизительно равен 1
- Закон Ципфа утверждает, что частота любого слова в публикации обратно пропорциональна его рангу в таблице частот. Таким образом, наиболее часто встречающееся слово будет встречаться примерно в два раза чаще, чем второе по частоте слово, в три раза чаще, чем третье по частоте слово, и т. д.



George Kingsley Zipf



Felix Auerbach



История инфометрии и библиометрии



Казанский Государственный
Медицинский Университет

INNOPOLIS
UNIVERSITY

ОПОРНЫЙ
ОБРАЗОВАТЕЛЬНЫЙ
ЦЕНТР

Закон Брэдфорда

- Впервые описан Сэмюэлем К. Брэдфордом в 1934 году. оценивает экспоненциально убывающую результаты поиска ссылок в научных журналах.
- Если журналы отсортировать в порядке убывания по количеству ссылок на опубликованные в них статьи и затем разделить журналы на три группы таким образом, чтобы каждая из групп журналов содержала бы примерно одну треть всех ссылок, то количество журналов в каждой группе будет пропорционально $1:n:n^2$.



Сэмюэл Клемент Брэдфорд (10 января 1878 г. в Лондоне - 13 ноября 1948 г.) был британским математиком, библиотекарем и документалистом из Музея науки в Лондоне.



История инфометрии и библиометрии



Казанский Государственный
Медицинский Университет

INNOPOLIS
UNIVERSITY

ОПОРНЫЙ
ОБРАЗОВАТЕЛЬНЫЙ
ЦЕНТР

Закон Прайса, Модель Прайса

- Закон квадратного корня Прайса или закон Прайса относится к взаимосвязи между литературой по теме и количеством авторов в предметной области, утверждая, что половина публикаций исходит от квадратного корня всех участников. Таким образом, если 100 статей написаны 25 авторами, пять авторов внесут 50 статей.
- Модель Прайса, математическая теория роста цитирования, основанная на том, что сейчас можно было бы назвать процессом предпочтительного прикрепления. Существующая статья получает новые цитирования пропорционально количеству существующих цитирований, которые уже есть у статьи.
- Исследования экспоненциального роста науки и периода «полураспада» научной литературы.



Дерек Джон де Солла Прайс (22 января 1922 - 3 сентября 1983) был британским физиком, историком науки и информатиком. Он был известен количественными исследованиями научных публикаций.



История инфометрии и библиометрии



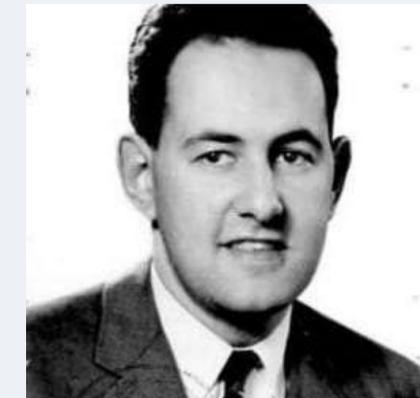
Казанский Государственный
Медицинский Университет

INNOPOLIS
UNIVERSITY

ОПОРНЫЙ
ОБРАЗОВАТЕЛЬНЫЙ
ЦЕНТР

Система информационного поиска SMART Джерарда Солтона

Солтон – «отец поиска информации», ведущий ученый-компьютерщик, работавший в области поиска информации, и. Его группа в Корнелле (Гарвард) разработала систему информационного поиска SMART для анализа и извлечения текста. Это была самая первая система, которая использовала популярную сейчас модель векторного пространства для поиска информации.



Джерард А.
«Джерри» Солтон (8
марта 1927 г. в
Нюрнберге - 28
августа 1995 г.)
профессор
компьютерных наук
Корнельского
университета.



История инфометрии и библиометрии



Казанский Государственный
Медицинский Университет

INNOPOLIS
UNIVERSITY

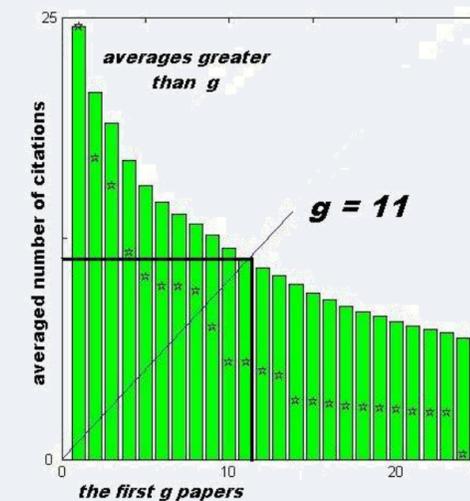
ОПОРНЫЙ
ОБРАЗОВАТЕЛЬНЫЙ
ЦЕНТР

g-индекс Лео Эгге

- *G*-индекс — это авторская метрика, предложенная в 2006 году Лео Эгге.
- *G*-индекс — наибольшее число n высокоцитируемых статей, ранжированных в порядке убывания количества цитирований, для которых среднее число цитирований составляет не менее n^2 . Таким образом, *g*-индекс 10 указывает на то, что 10 лучших публикаций автора цитировались не менее 100 раз (10^2), *g*-индекс 20 указывает на то, что 20 лучших публикаций автора цитировались 400 раз (20^2).



Leo Egghe



История инфометрии и библиометрии



Казанский Государственный
Медицинский Университет

INNOPOLIS
UNIVERSITY

ОПОРНЫЙ
ОБРАЗОВАТЕЛЬНЫЙ
ЦЕНТР

g-индекс Лео Эгге

- *G*-индекс является альтернативой предложенному ранее *h*-индексу Хирша. Хирш-индекс не усредняет количество цитирований, вместо этого он определяет минимум *n* цитирований для наименее цитируемой статьи в наборе и, таким образом, игнорирует количество цитирований очень часто цитируемых статей. *h* — это количество статей с порогом качества, которое растет по мере роста *h*; *g* позволяет использовать цитаты из статей с более высоким цитированием для поддержки статей с более низким цитированием в соответствии с этим порогом. По сути, *g*-индекс является максимально достижимым значением *h*-индекса, если фиксированное количество цитирований может быть свободно распределено по фиксированному количеству статей. Следовательно, во всех случаях *g* не меньше *h*, а в большинстве случаев выше. *G*-индекс часто разделяет авторов на основе цитирований в большей степени, чем *h*-индекс.
- *g*-индекс насыщается всякий раз, когда среднее количество цитирований всех опубликованных статей превышает общее количество опубликованных статей; если автор с насыщенным *g*-индексом опубликует больше статей, его *g*-индекс увеличится.



История инфометрии и библиометрии



Казанский Государственный
Медицинский Университет

INNOPOLIS
UNIVERSITY

ОПОРНЫЙ
ОБРАЗОВАТЕЛЬНЫЙ
ЦЕНТР

Рональд Руссо

Рональд Руссо (1949, Антверпен) — бельгийский математик и информатик. Он получил международную репутацию за свои исследования индикаторов и анализа цитирования в области библиометрии и наукометрии, разработке и использовании показателей для измерения качества исследований и основных тенденций в науке. Эксперт в области анализа цитирования и оценки исследований.

В 1990 году Рональд Руссо и Лео Эгге написали первое руководство по библиометрии. Руссо опубликовал более 200 научных статей по различным аспектам библиометрии и наукометрии, в том числе в журналах *Scientometrics*, *Journal of the American Society for Information Science and Technology* и *Journal of Informetrics*. Он считается одним из самых продуктивных писателей в этой дисциплине. Руссо вместе с Лео Эгге получил премию Дерекка де Солла Прайса в 2001 году. С 2007 по 2015 год он был седьмым председателем ISSI. В 2015 году он стал соредактором журнала *Data and Information Science*.



Рональд Джозеф
Эмиль Руссо
14 августа 1949
г.
Антверпен
(Бельгия)



История инфометрии и библиометрии



Казанский Государственный
Медицинский Университет

INNOPOLIS
UNIVERSITY

ОПОРНЫЙ
ОБРАЗОВАТЕЛЬНЫЙ
ЦЕНТР

Манфред Бониц

Бониц принадлежит к пионерам информационных наук в социалистических странах и посвятил значительную часть своей научной жизни только что возникшей области наукометрии. Его членство в редколлегии журнала «Наукометрия» продлилось более 20 лет. В этом журнале появилось большинство его библиометрических публикаций.



Манфред
Бониц
(7.3.1931 -
14.08.2012)
немецкий
информатик и
физик



История инфометрии и библиометрии



Казанский Государственный
Медицинский Университет

INNOPOLIS
UNIVERSITY

ОПОРНЫЙ
ОБРАЗОВАТЕЛЬНЫЙ
ЦЕНТР

Набор инструментов VibExcel Олле Перссон

Олле Перссон предложил VibExcel для анализа библиографических данных или любых данных текстового характера, отформатированных аналогичным образом. Идея состоит в том, чтобы создавать файлы данных, которые можно импортировать в Excel или любую программу, которая берет записи данных с вкладками, для дальнейшего использования и обработки.



Олле Перссон



История инфометрии и библиометрии



Казанский Государственный
Медицинский Университет

INNOPOLIS
UNIVERSITY

ОПОРНЫЙ
ОБРАЗОВАТЕЛЬНЫЙ
ЦЕНТР

Библиометрия

- Библиометрия — это использование статистических методов для анализа книг, статей и других публикаций, особенно в отношении научного содержания
- Библиометрические методы часто используются в области библиотек и информатики
- Библиометрия тесно связана с наукометрией, обе области во многом пересекаются



История инфометрии и библиометрии



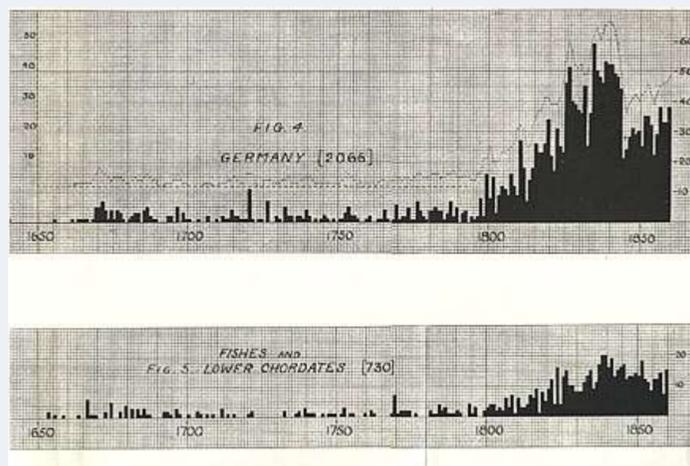
Казанский Государственный
Медицинский Университет

INNOPOLIS
UNIVERSITY

ОПОРНЫЙ
ОБРАЗОВАТЕЛЬНЫЙ
ЦЕНТР

Библиографическая статистика

- 1917 г. Francis Joseph Cole и Nellie B. Eales: статистическая ценность публикаций может быть оценена
- 1922 г. Edward Wyndham Hulme: публикации можно рассматривать как стандартную меру всей цивилизации



Ранний пример библиометрического анализа научного корпуса по анатомии Фрэнсиса Джозефа Коула и Нелли Б. Илз в 1917 году с разбивкой по темам и странам

<https://en.wikipedia.org/wiki/Bibliometrics>



История инфометрии и библиометрии



Казанский Государственный
Медицинский Университет

INNOPOLIS
UNIVERSITY

ОПОРНЫЙ
ОБРАЗОВАТЕЛЬНЫЙ
ЦЕНТР

Индексы цитирования

- 1927 г. P. Gross и E. M. Gross ранжировали журналы по химии в зависимости от уровня их цитирования, предложили присвоение бонуса недавним цитированиям
- Alfred Lotka закон продуктивности на основе анализа авторских публикаций
- Samuel Bradford закон рассеяния - результаты поиска ссылок в научных журналах экспоненциально уменьшаются



История инфометрии и библиометрии



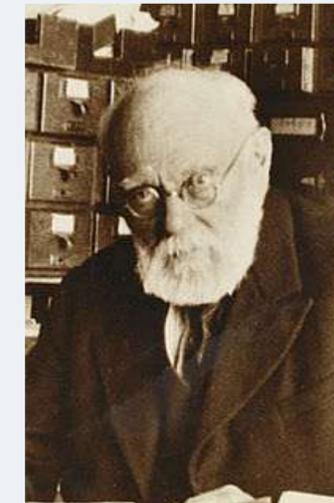
Казанский Государственный
Медицинский Университет

INNOPOLIS
UNIVERSITY

ОПОРНЫЙ
ОБРАЗОВАТЕЛЬНЫЙ
ЦЕНТР

Библиометрия или статистическая библиография?

- Термин *bibliométrie* предложил Paul Otlet в 1934 году – «измерение всех аспектов, связанных с публикацией и чтением книг и документов»
- 1939 г John Desmond Bernal предсказал развитие научных архивов
- Англизированная версия библиометрии впервые использована Alan Pritchard в 1969 году – «применение математических и статистических методов к книгам и другим средствам коммуникации»



Paul Marie
Ghislain Otlet
23/08/1868 –
10/12/1944)



История инфометрии и библиометрии



Казанский Государственный
Медицинский Университет

INNOPOLIS
UNIVERSITY

ОПОРНЫЙ
ОБРАЗОВАТЕЛЬНЫЙ
ЦЕНТР

Периодический кризис, оцифровка и индекс цитирования

1953 г. – первый симпозиум по машинному индексированию, профессор Л. Рид (L. Reed – вице-президент Университета Дж. Гопкинса): «Человек тонет в потоке информации» [Hyslop, 1953].



Профессор Л. Рид



История инфометрии и библиометрии



Казанский Государственный
Медицинский Университет

INNOPOLIS
UNIVERSITY

ОПОРНЫЙ
ОБРАЗОВАТЕЛЬНЫЙ
ЦЕНТР

Компьютерный поиск результатов исследований

- 1957 г. Ганс Петер Лун разработал парадигму статистического анализа частотности слов (KWIC – key word in context)
- Создание «централизованной информационной системы» SCITEL. Автоматизированный перевод неанглоязычной научной литературы
- Выборочное распространение информации («SDI»)



Hans Peter Luhn – создатель алгоритма Луна, индексации KWIC (Key Words In Context) и выборочного распространения информации («SDI»).

История инфометрии и библиометрии



Казанский Государственный
Медицинский Университет

INNOPOLIS
UNIVERSITY

ОПОРНЫЙ
ОБРАЗОВАТЕЛЬНЫЙ
ЦЕНТР

Инновационные библиографические продукты

- В 1956 году Гарфилд основал Институт научной информации, который стал основной частью научного подразделения Thomson Reuters, теперь известного как Clarivate Analytics.
- Инновационные библиографические продукты:
 - Current Contents,
 - Science Citation Index (SCI) и другие базы данных цитирования,
 - Journal Citation Reports (JCR)
 - Index Chemicus.
- Редактор-основатель и издатель The Scientist, новостного журнала для ученых-биологов
- Алгоритм HITS и PageRank - структурированное цитирование между сайтами через гиперссылки, лежащие в основе поисковой системы Google



Eugene Eli Garfield
(16.09.1925 -
26.02.2017)
Американский
лингвист и
бизнесмен, один из
основоположников
библиометрии и
наукометрии.



История инфометрии и библиометрии



Казанский Государственный
Медицинский Университет

INNOPOLIS
UNIVERSITY

ОПОРНЫЙ
ОБРАЗОВАТЕЛЬНЫЙ
ЦЕНТР

Эра Интернет

- 1997 г. — SCI, SSCI, ANCI, CPCI-S и CPCI-SSH, а также JCR-Science Edition, JCR-Social Science Edition были объединены в единую политематическую реферативную базу данных научных публикаций с указанием цитирования — Web of Science Core Collection.
- 2008–2010 гг. — расширена география научных журналов, индексируемых в Web of Science Core Collection. В настоящее время в базе данных индексируются приблизительно двенадцать с половиной тысяч лучших международных научных журналов по всем отраслям знания.



История инфометрии и библиометрии



Казанский Государственный
Медицинский Университет

INNOPOLIS
UNIVERSITY

ОПОРНЫЙ
ОБРАЗОВАТЕЛЬНЫЙ
ЦЕНТР

Библиометрические показатели

- количество публикаций
- их цитируемость
- импакт-фактор научного журнала, в котором они были опубликованы
- количество высокоцитируемых статей
- индекс Хирша
- различные нормализованные показатели цитируемости

Дополнительно:

- количество грантов, научных премий,
- участия в составах редколлегии ведущих научных журналов.

Руководство по наукометрии: индикаторы развития науки и технологии : [монография] / М. А. Акоев, В. А. Маркусова, О. В. Москалева, В. В. Писляков ; [под. ред. М. А. Акоева]. – Екатеринбург : Изд-во Урал. ун-та, 2014. – 250 с.



INNOPOLIS
UNIVERSITY

● ОПОРНЫЙ
ОБРАЗОВАТЕЛЬНЫЙ
ЦЕНТР



Казанский Государственный
Медицинский Университет

Спасибо
за внимание