

Monsoon 2020

2 - Basics of Information Retrieval

Information Retrieval

by

Dr. Rajendra Prasath



Indian Institute of Information Technology

Sri City – 517 646, Andhra Pradesh, India

Information Retrieval?

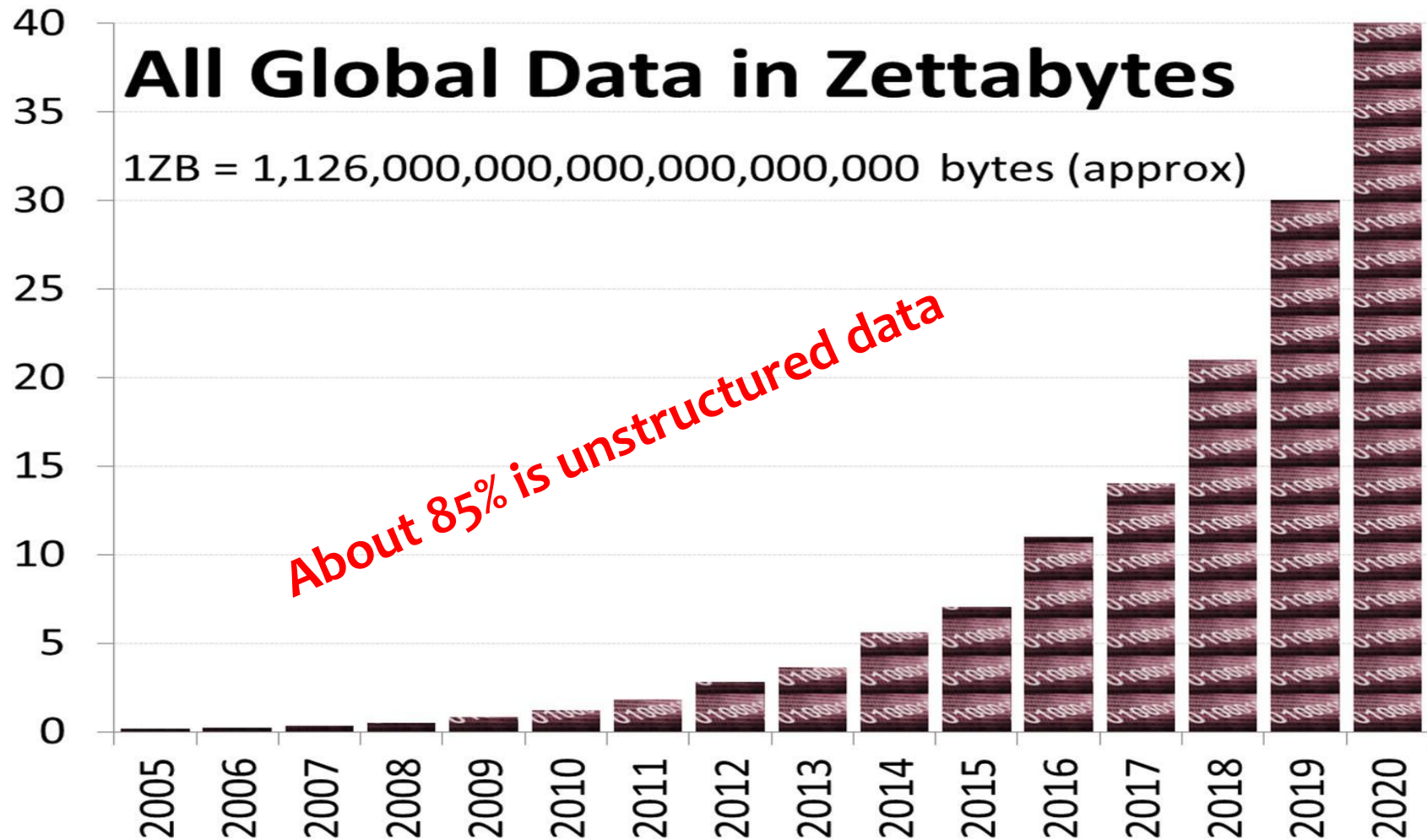


Information Retrieval

- **Information Retrieval (IR)** is finding material (usually documents) of an unstructured nature (usually text) that satisfies an information need from within large collections (usually stored on computers).
- These days we frequently think first of web search, but there are many other cases:
 - E-mail search
 - Searching your laptop
 - Corporate knowledge bases
 - Legal information retrieval
 - and so on . . .

Size – Scaling of Text data

Huge Data ☐ mind – blowing size (1ZB = 10^{21} bytes)



Data

How do we define Data?

- Look at the following data:

80.32	80.32	91.54
76.23	76.23	80.32
54.24	54.24	76.23
64.96	64.96	64.96
91.54	91.54	54.24

- What do these numbers refer to?
 - **Marks** or **Heights** or **Weights** or . . .

Structured Data

- Look at the following data:

Export Item	13 – 17 Jan 2018*	Usage	13 – 17 Jan 2018*
Turmeric	80.32	Turmeric	80.32
Turmeric	76.23	Turmeric	76.23
Turmeric	54.24	Turmeric	54.24
Turmeric	64.96	Turmeric	64.96
Turmeric	91.54	Turmeric	91.54

* In metric tons

* Average usage of a nuclear family in Hyd (in milligrams)

- Does it make any sense now?

Structured Data (contd)

- Let us look into the details

		Name	Weight	Name	2018 (25 Dec)
Rahul	80.32	Rohan	91.54	Rohan	91.54
Priya	76.23	Priya	80.32	Priya	80.32
Sowmya	54.24	Rahul	76.23	Rahul	76.23
Lucky Vogel	64.96	Lucky Vogel	64.96	Lucky Vogel	64.96
Rohan	91.54	Sowmya	54.24	Sowmya	54.24

- More Data ... More Features ... More Value(s)
- More Value ☐ More Money (?!)

Text Data

No Free Meals: Transaction Fees are not the same
13 Jan 2018

Mumbai: Manhattan FinServices Inc. has analyzed mobile payment products. As per RBI analytics, find that includes the highest and lowest transaction products. **PayGO** ranks top in the list with **91.54** paise charging the lowest transaction fee of **54.24** paise. Amazon charges **64.96** paise per transaction for the transaction fee of **Paypal** and **Paytm** are given as observed that **PayPaise** is charging nearly half of **PayGO**.

Many interesting facts revealed in this report would be interested to opt for the specific payment service. With mess transaction fee.

Product	13 Jan 2018 *
Paytm	80.32
PayPal	76.23
Rupay	54.24
Amazon Pay	64.96
PayGO	91.54

* Fee in Paise(INR) per transaction
in Mumbai

- Does the table give a better understanding now?

What do you understand?

Non-Banking Financial Companies (NBFC) are establishments that provide financial services and banking facilities without meeting the legal definition of a Bank. The top 10 companies are listed by The Banking Finance Post on November 17, 2018 says that Power Finance Corporation Limited ranks topper with INR 267377.40 in terms of annual turn over which is measure in millions of rupees. The second and third places are held by Rural Electrification Corporation Limited (INR 224403.10 millions) and Bajaj Finance Limited (INR 133292.20 millions). The remaining companies in the top 12 list are given as follows with their annual turnover in the brackets respectively: Shriram Transport Finance Company Limited (122768.30), Indian Railway Finance Corporation Limited (110202.32) Mahindra & Mahindra Financial Services Limited (72061.20) HDB Financial Services Limited (70619.90) Muthoot Finance Limited (62432.00) Cholamandalam Investment and Finance Company Limited ((54257.60) L&T Finance Limited (52460.00) Shriram City Union Finance Limited (51015.70), and Tata Capital Financial Services Limited (45553.70).

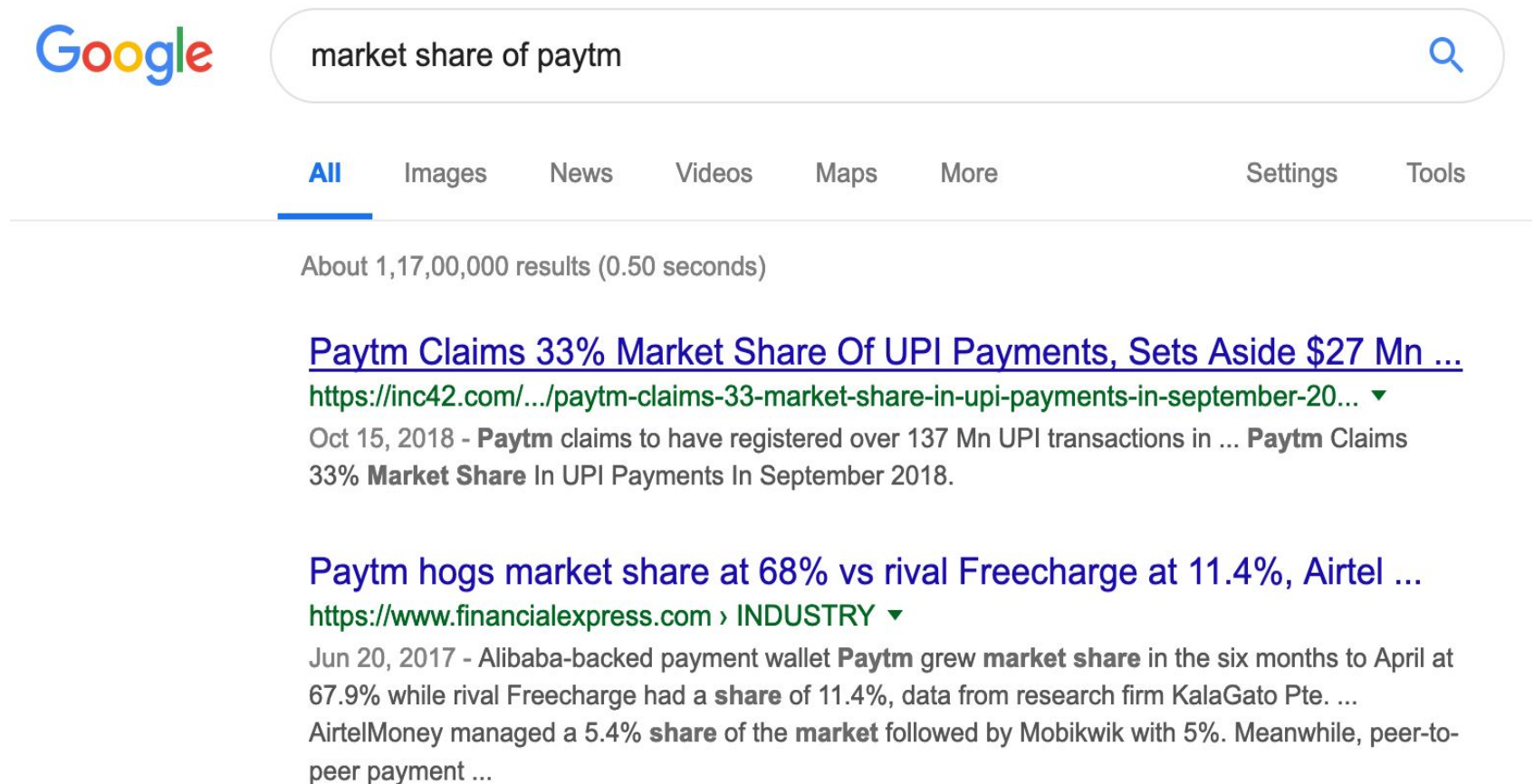
Try with this!!

Top 50 NBFCs' Ranking Based on Annual Turnover

NBFCs List	Total Income	Rank
Power Finance Corporation Limited	267377.40	1
Rural Electrification Corporation Limited	224403.10	2
Bajaj Finance Limited	133292.20	3
Shriram Transport Finance Company Limited	122768.30	4
Indian Railway Finance Corporation Limited	110202.32	5
Mahindra & Mahindra Financial Services Limited	72061.20	6
HDB Financial Services Limited	70619.90	7
Muthoot Finance Limited	62432.00	8
Cholamandalam Investment and Finance Company Limited	54257.60	9
L&T Finance Limited (erstwhile Family Credit Limited)	52460.00	10
Shriram City Union Finance Limited	51015.70	11
Tata Capital Financial Services Limited	45553.70	12

Words, Relevance & Personalization

✧ Market Share of **paytm**



The screenshot shows a Google search interface. The search bar contains the text "market share of paytm". Below the search bar, the "All" tab is selected. The search results show approximately 1,17,00,000 results in 0.50 seconds. The first result is a link to "Paytm Claims 33% Market Share Of UPI Payments, Sets Aside \$27 Mn ..." with a URL starting with "https://inc42.com/...". The second result is a link to "Paytm hogs market share at 68% vs rival Freecharge at 11.4%, Airtel ..." with a URL starting with "https://www.financialexpress.com > INDUSTRY".

Google

market share of paytm

All Images News Videos Maps More Settings Tools

About 1,17,00,000 results (0.50 seconds)

[Paytm Claims 33% Market Share Of UPI Payments, Sets Aside \\$27 Mn ...](#)
<https://inc42.com/.../paytm-claims-33-market-share-in-upi-payments-in-september-20...> ▼
Oct 15, 2018 - **Paytm** claims to have registered over 137 Mn UPI transactions in ... **Paytm** Claims 33% **Market Share** In UPI Payments In September 2018.

[Paytm hogs market share at 68% vs rival Freecharge at 11.4%, Airtel ...](#)
<https://www.financialexpress.com > INDUSTRY> ▼
Jun 20, 2017 - Alibaba-backed payment wallet **Paytm** grew **market share** in the six months to April at 67.9% while rival Freecharge had a **share** of 11.4%, data from research firm KalaGato Pte. ... AirtelMoney managed a 5.4% **share** of the **market** followed by Mobikwik with 5%. Meanwhile, peer-to-peer payment ...

Words and Knowledge Bases

✧ Make your judgments on relevance

Paytm registers 600% growth in UPI transactions in 6 months

<https://www.hindustantimes.com/.../paytm.../story-iLTPwMRjKqj55fNCVnR5QK>

Nov 2, 2018 - **Paytm** on Friday said it witnessed 600 % growth in the Unified Payment Interface (UPI) payments with over 33 per cent of the overall **market share**. ... With t over 80 % share of all offline merchant

Paytm hogs market share at 68% vs rival Freecharge a

<https://www.financialexpress.com> > INDUSTRY ▼

PhonePe becomes largest UPI player, leaps ah

<https://www.digit.in> > Apps ▼

Aug 2, 2018 - Of these PhonePe had the largest **market share** of network, ahead of other payment apps like **Paytm** and Tez.

Berkshire Hathaway Takes Stake In India's Pay

<https://www.forbes.com/sites/.../08/.../berkshire-hathaway-tak>

Aug 27, 2018 - Berkshire Hathaway has taken a stake in **Paytm**, the lion's **share** of the **market** with 9.9%, followed by PayPal ...

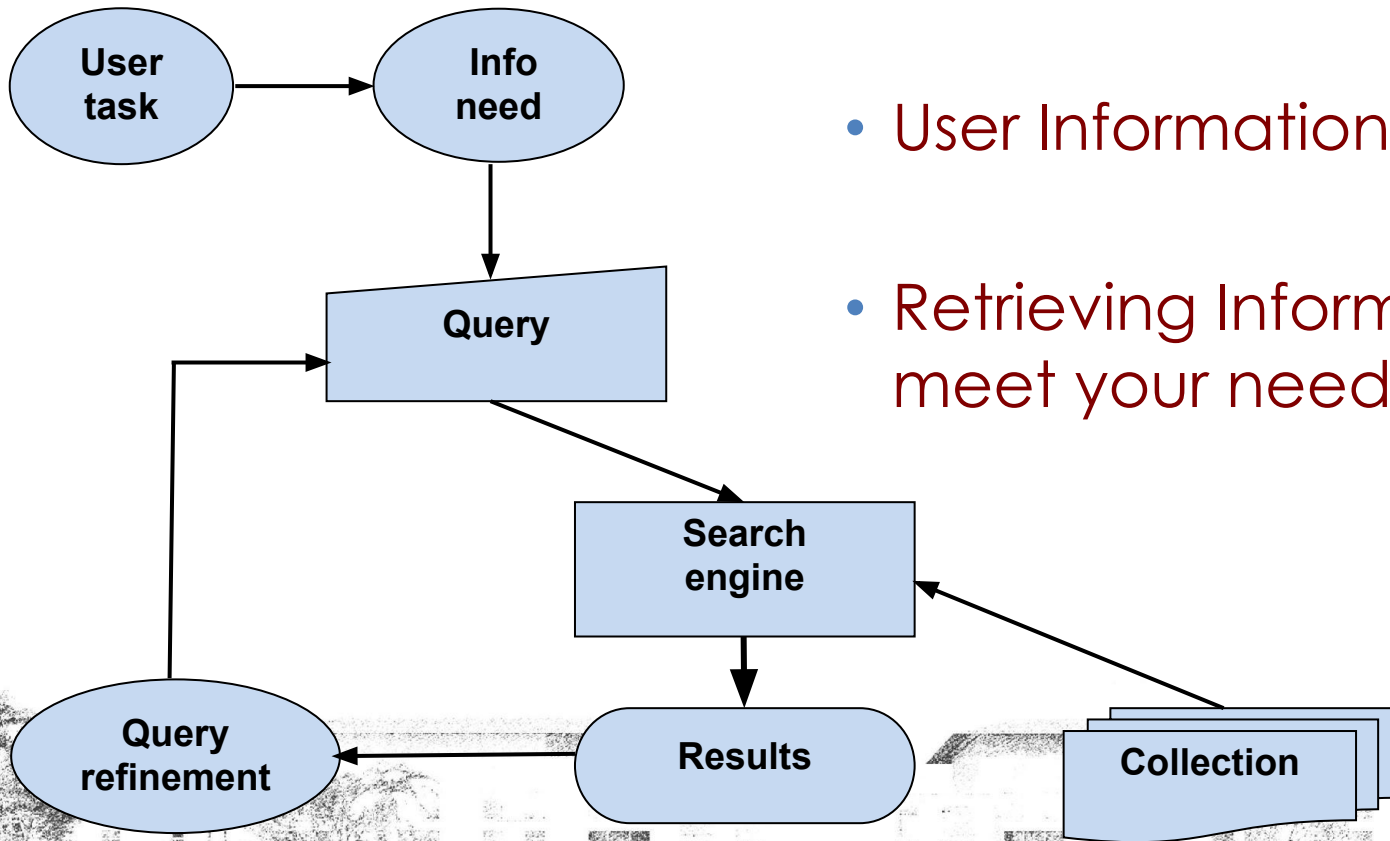
paytm

Type of business	Private
Type of site	E-commerce
Founded	2010
Headquarters	Noida, Uttar Pradesh, India
Area served	India, Canada
Founder(s)	Vijay Shekhar Sharma
Key people	Vijay Shekhar Sharma (CEO)
Industry	Internet
Products	Paytm Mall Paytm Payments Bank Paytm Money Paytm Gamepind
Services	Online shopping, payment systems, digital wallets
Revenue	₹814 crore (US\$110 million) (FY 2017) ^[1]
Parent	One97 Communications Ltd
Website	paytm.com 

Classical Search Engines

Bridging the gap:

- User Information Needs
- Retrieving Information to meet your needs



Text Collections

✧ Structured data

- ✧ Information stored DB
- ✧ Strict format
- ✧ Limitation
- ✧ Not all data collected is structured

✧ Semi-structured data

- ✧ Data may have certain structure but not all information collected has identical structure
- ✧ Some attributes may exist in some of the entities of a particular type but not in others

✧ Unstructured data

- ✧ Very limited indication of data type
- ✧ For example, look into a simple text document

Words and their occurrences

- ✧ Consider any city:
- ✧ Check with Wikipedia
 - ✧ For e.g, search for **Darjeeling** (English Version)
 - ✧ <https://en.wikipedia.org/wiki/Darjeeling>
- ✧ How many times Darjeeling comes up in the document?
- ✧ Does it mean the following?
 - ✧ more it occurs ... more it is important?

Look at Text Segment

- ❖ **Darjeeling** is a city and a municipality in the Indian state of West Bengal. It is located in the Lesser Himalayas at an elevation of 6,700 ft. It is noted for its tea industry, its views of Kangchenjunga, the world's third-highest mountain, and the **Darjeeling** Himalayan Railway, a UNESCO World Heritage Site. **Darjeeling** is the headquarters of the **Darjeeling** District which has a partially autonomous status within the state of West Bengal. It is also a popular tourist destination in India.

Words – Counts

✧ A subset of words and their counts

7 the

5 of

5 is

5 a

4 Darjeeling

3 in

3 It

2 state

2 its

2 and

2 West

2 Bengal.

1 world's

1 within

1 which

1 views

1 tourist

.

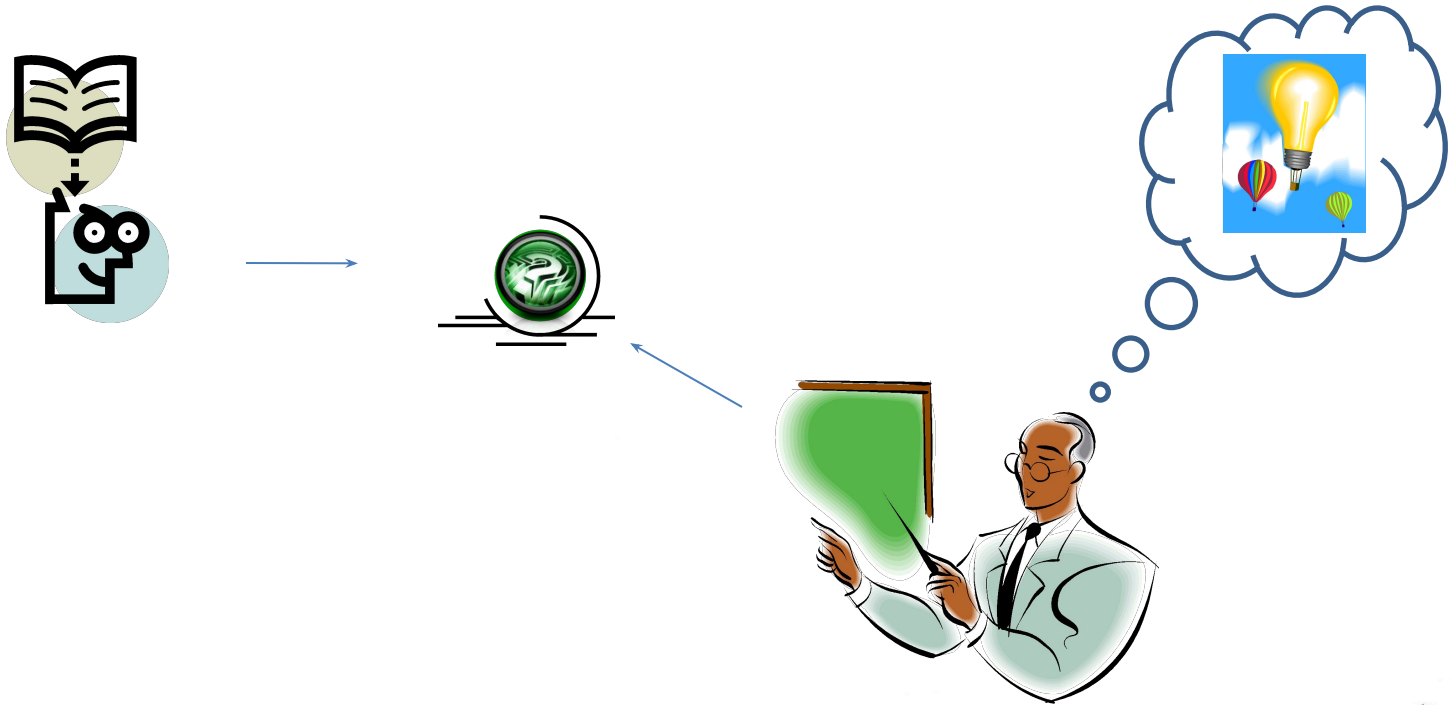
.

.

Assistance

- ✧ You may post your questions to me at any time
- ✧ You may meet me in person on available time or with an appointment
- ✧ You may leave me an email any time (email is the best way to reach me faster)

Thanks ...



... Questions ???