#1

Complete the following sentences:

1- …. is a system of hypertext documents that are viewable by a web browser. Runs as service/application on top of internet protocols.

(World wide web) <- This is the answer.

2- …. are documents with internal cross-references to other documents.

(Hypertext documents)

3- Web based on ……, …… and …… As fundamental mechanisms.

(Application protocol, Document description format and document addressing mechanism)

4- ….. is a buzzword for describing modern forms of applications and services that are accessible by web browser such as wikis or weblogs.

(Web 2.0)

5- …… is a vision of semantically interconnected data in the web.

(Web 3.0/Semantic web)

6- XHTML stands for .....

(Extensible HyperText Markup Language)

7- DTD stands for .....

(Document Type Definition)

8- DOM stands for ....

(Document Object Model)

9- …… is a client-side language for defining layout of mark-up language elements. It simplifies uniform design and its maintenance of web pages.

(CSS: Cascading Style Sheets)

10- ……. Is a general purpose meta-language for representation, exchange and processing of data.

(XML)

11- In XML, all attribute values are ….. .

(single or double quoted)

12- Syntactically correct XML documents are called ……. .

(Well-Formed)

13- If different contributors use the same element names to refer to different things in one file, …… will occur.

(name conflicts)

14- Name conflict are solved by ….. which are used to qualify elements and attributes by unique identifiers.

(XML namespaces)

15- A …… XML document is related to and conforms to further information describing its structure and data types using therein.

(Valid) .. yeah no kidding just valid.

16- …… and …… are two often used type definition languages.

(DTD and XML Schema)

17- …… is part of the family of XML markup languages. It mirrors or

extends versions of the widely used Hypertext Markup Language (HTML).

(XHTML)

18- ..... is an application programming interface standardized by the W3C for accessing and manipulating XML documents.

(DOM)

19- ..... interprets the document in its logical tree structure thus making it necessary to load a tree representation of the document into memory.

(DOM)

20- …… is an event-based mechanism for XML processing.

(SAX/simple API for XML)

21- …… Compact text-based and human readable data serialization format, Superset of JSON, Offers lists, associative arrays and scalars for structuring data and Enables cross-references between nodes.

(YAML: Ain’t Markup Language)

22- In order to find an appropriate interpreter program for received content on the client side the

content's meta information specifies a ……. .

(Content Type/media type/MIME type)

23- ……. is an object based wide-spread script language often used for client-side execution.

(Javascript)

#2

1- ….. is a software that makes resources available through an interface that is accessible by HTTP.

(Web Server)

2- Current web servers are structured in a hierarchical way that allows ……, due to the need for ……. .

(parallel request processing/ low latency time)

3- On start-up, …… creates a pool of entities for client handling and opens an interface for incoming connection requests.

(Web Server)

4- …….. initiates communication on the basis of TCP through this interface which is handled by a connector entity.

(Web browser)

5- ……. Creates a client descriptor which is the endpoint of a communication path to the client and passes it to the delegation process.

(Connector entity)

6- …….. chooses an available pregenerated processing entity and

forwards the client descriptor to it.

(delegation process)

7- …… today are extended by technologies as XML support, script engines and plugin engines.

(Web browsers)

8- …… is the front-end for displaying a page to the user of the web browser.

(User-interface)

9- ……. embedded browser component that provides a high level for querying and using the rendering engine.

(Browser engine)

10- ……. Performs parsing and layout for XHTML documents enriched with other languages as CSS in the browser.

(Rendering engine)

11- ……. realizes HTTP communication of the browser with the server.

(Networking)

12- …….. executes scripts embedded in XHTML pages in the web browser.

(Script engine)

13- ..... provides drawing and windowing primitives, user interface widgets and fonts for the web browser.

(Display backend).

14- ..... stores web browser associated data.

(data persistence)

15- ..... is a **technique** that makes it possible to send requests to the server without the need to wait for the response of previous requests.

(pipelining)

16- …….. is a part of the HTTP1.1 request message that specifies the method to be performed, the associated web resources and the HTTP version.

(Request line)

17- …….. is a part of the HTTP1.1 request message that gives general information such as the type of connection or the data of request generation.

(General Header)

18- …….. allows to pass information about the HTTP1.1 request message and about the client to the server.

(Request header)

19- …….. is a part of the HTTP1.1 response message that specifies the HTTP protocol version and status code of request processing in numeric and textual representation.

(status line)

20- The code that started with……. means for example that the client should continue sending the request.

(1xx)

21- The code that started with……. means for example that request has

succeeded.

(2xx)

22- The code that started with…….. means for example that requested resource has been moved permanently to new URI.

(3xx)

23- The code that started with…….. means for example that server has not found anything matching the request-URI.

(4xx)

24- The code that started with……. means for example that server doesn't support the functionality required to fulfil the request.

(5xx)

25- ……. Provide an optimized transport for HTTP semantics using …….. and ……. .

(HTTP/2 … multiplexing …. Header compression)

#3

1- …… means that in order to update the displayed data, an HTTP request is sent and the user actions are blocked until an HTTP response from the server was sent.

(Synchronous communication model)

2- ……. requires high degree of dependence between HTTP request and response. For every data transfer from server to client, a request has to be sent directly before this transfer.

(Synchronous communication model)

3- …….. is the coupling of user activity and server-side processing.

(Synchronous communication model)

4- …… is the click, wait and refresh user interaction paradigm.

(Synchronous communication model)

5- …….. is a model in which the browser reacts to a user action by sending an HTTP request to the web server then the server processes the request and sends a response with a complete XHTML page to the browser.

(Synchronous communication model)

6- ……. Is a model that the browser refreshes the screen and displays a new XHTML page after each user action.

(Synchronous communication model)

7- One of the shortcoming of the classic web application model is …….. .

-> The shortcomings are:
(Slow Performance
&&
loss of operation context
&&
excessive server load and bandwidth consumption
&&
Unreliability and inefficient
&&
low productivity and interactivity)

8- ….. is creating web application that look like desktop application using several technologies, each flourishing in its own right , coming together in powerful new ways.

(AJAX) :D .. anyways..

9- ……. Eliminates the click, wait and refresh user interaction paradigm on web by introducing an intermediary called …….. .

(AJAX/AJAX engine)

10- …… responsible for rendering the user interface and communicating with the server on the user's behalf.

(AJAX engine)

11- …… allows the user's interaction with the application in an asynchronous way.

(AJAX engine)

12- User actions generates a …… to the AJAX engine instead of …… .

(Javascript call / HTTP Request)

13- AJAX engine sends ……. To the server to get needed data from the server.

(Asynchronous HTTP Request)

14- …….. updates the displayed data without blocking user actions by waiting for the HTTP response from the server.

(Non blocking communication model)

15- ……. requires high degree of dependence between HTTP request and HTTP response, for each server-side processes update, a full request/response round trip between AJAX engine and server is necessary.

(Non blocking communication model)

16- ……. Is the model that user activity and server-side processing are decoupled.

(Non blocking communication model)

17- JSON stands for …… .

(Javascript Object Notation)

18- …….. enables web servers to send data to the client with only an initial HTTP request and after this without having any need for the client to request the data.

(Comet/Asynchronous communication model)

19- Comet stands for …… .

(nothing, just a buzzword) xD

20- ……. is used for applications in which available data should be transmitted in real-time.

(Comet)

21- HTTP-based streaming central shortcomings are ……., …….. .

(overhead and communication channel is unidirectional(inefficiency))

22- ..... is a standard for TCP-based bidirectional communication channels established using the HTTP protocol avoiding the drawback of long-polling.

(WebSockets)

#5

1- ……. carry out the collection and categorization of data in the search engine.

(Web Crawler/robot/scutter)

2- ……. Deployed in many application domains where data extraction is necessary.

(Web Crawler/robot/scutter)

3- fetching of pages is realized by ……. .

(regular HTTP client.)

4- ……. Crawling strategy Always follow all links to adjacent pages of the current page first.

(Breadth-first)

5- ……. Crawling strategy adds all found URIs of adjacent pages to a queue.

(Breadth-first)

6- If all links of the current page have been visited, ……. Crawling strategy takes the first page of its queue and start over.

(Breadth-first)

7- …….. strategy in crawling always follow the first unvisited link of the current page

(Depth-first)

8- ……. declares this visited page as the current page and start over.

(Depth-first)

9- ……realize backtracking at dead ends or after reaching a predefined depth.

(Depth-first)

10- ..... URI crawler-dependent representation.

(uniform canonical/normalized)

11- Access to unstructured web data is called ..... .

(Web scraping)

12- ..... is a major problem that increase difficulties to access content while web scraping.

(Not well-formed HTML)

13- ..... Is the transformation of badly written XHTML pages to well

formed one.

(Tidying)

14- ..... is considered as a representation of comfortable unstructured information accessing within XHTML.

(DOM)

15- ..... is a standardized way that allows to inform web crawlers about available URIs and some of their properties on a website.

(Sitemaps protocol)

16- ..... is a function of query frequency within a document (TF) and across all documents (IDF).

(Ranking)

17- Central assumptions that count for classical document collections cannot be applied to the web because ..... .

(Short web queries
OR
enormous variety of documents
OR
misinformation on many pages)

18- ..... Can be seen as document relevance and popularity.

(Linkage intensity)

19- ……. is a way to prioritize results of web keyword search based on evaluation of the link structure.

(PageRanking)

20- …… invented by google founders and applied as part of google's ranking algorithm

(PageRanking)

21- Hyperlink from page A to page B is a ……. Of the content of page B by author of page A

(Recommendation)

22- Quality of page is related to …… which means number of incoming links.

(its in-degree)

23- Quality of a page is related to …… and ……. .

(its in-degree / quality of pages linking to it.)

24- HITS stands for ……. .

(Hyperlinked Induced Topic Search)

25- ..... differentiates between Hubs and Authorities and delivers result lists for both categories.

(HITS)

26- ..... is a page with many in-links.

(Authority)

27- ..... is a page with many out-links.

(Hub)

28- Authority comes from ....., while hub comes from ..... .

(in-edges / out-edges)

Bonus Round

1- …….. is a web of data that machines can understand too.

(Semantic Web)

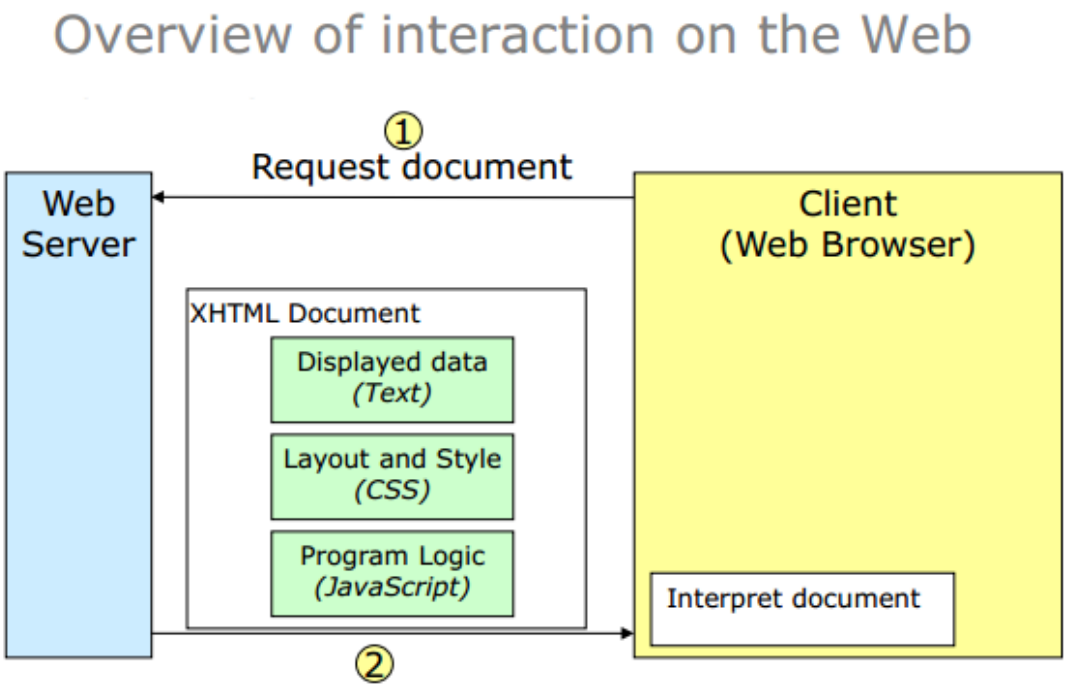
2- ……. Is the new standard for HTML, designed to deliver almost everything you want to do online without requiring additional plugins. Also its cross-platform.

(HTML5)

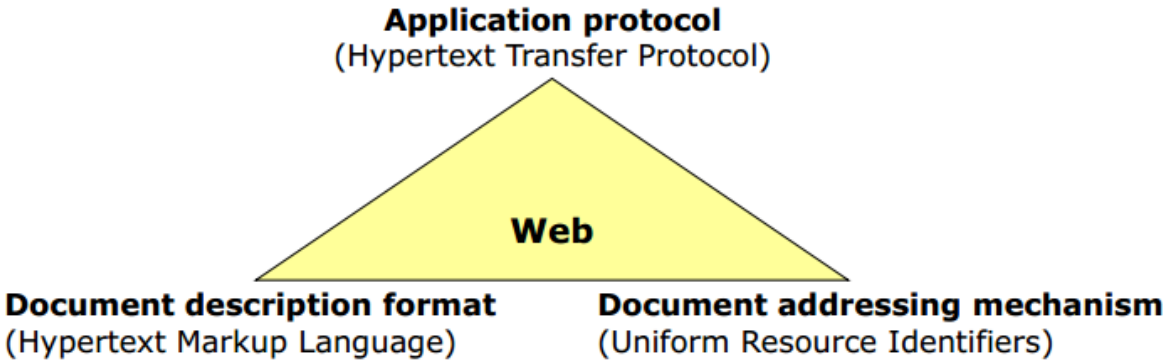
3- Yeah there are still a lot of stuff in both presentations. GO check them out again ftw.

<<---The 100 Questions--->>

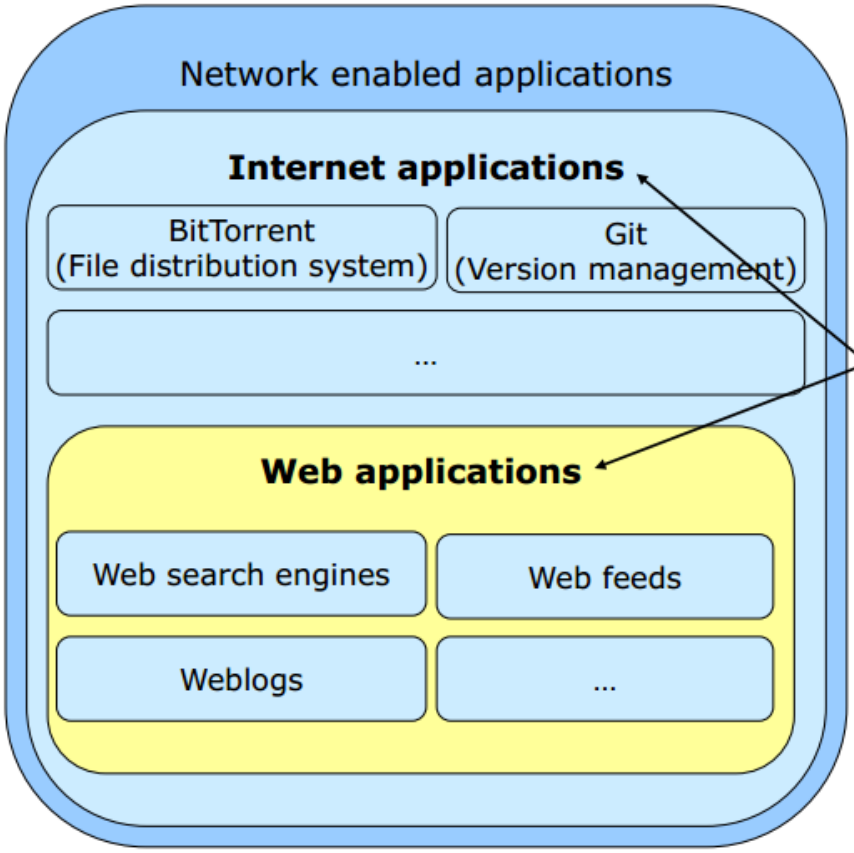
Some important and not important drawings ! …



Web based on:



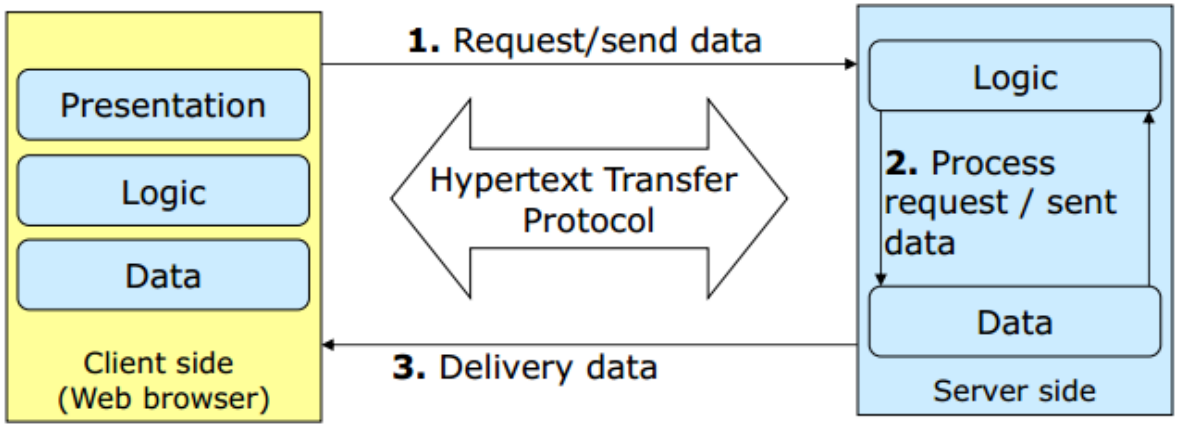
This is not that important ya know.

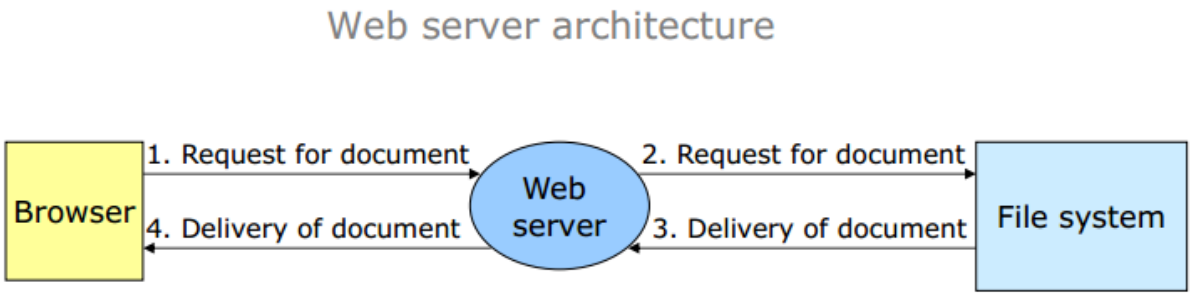


Subset of important
- applications,
- mechanisms,
- protocols and
- languages

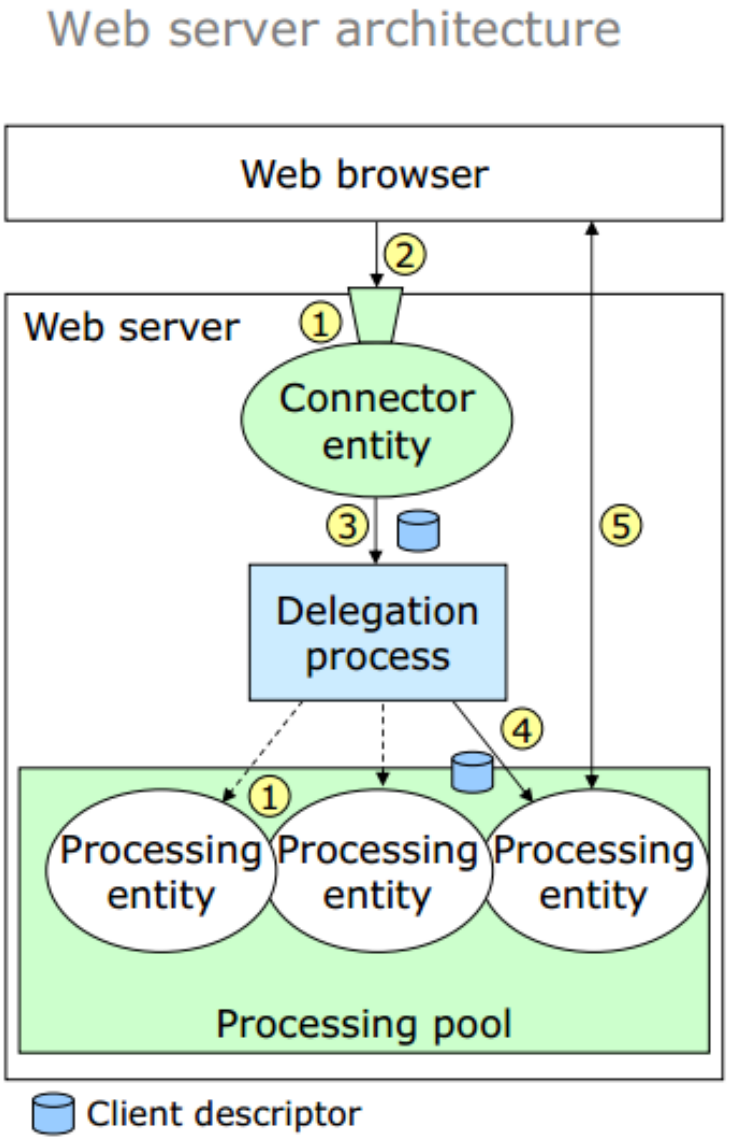
that are associated with these two categories will be discussed in the lecture

slide 2

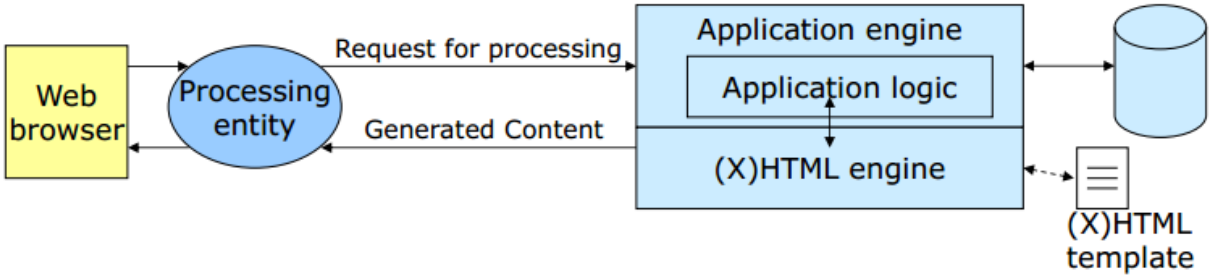


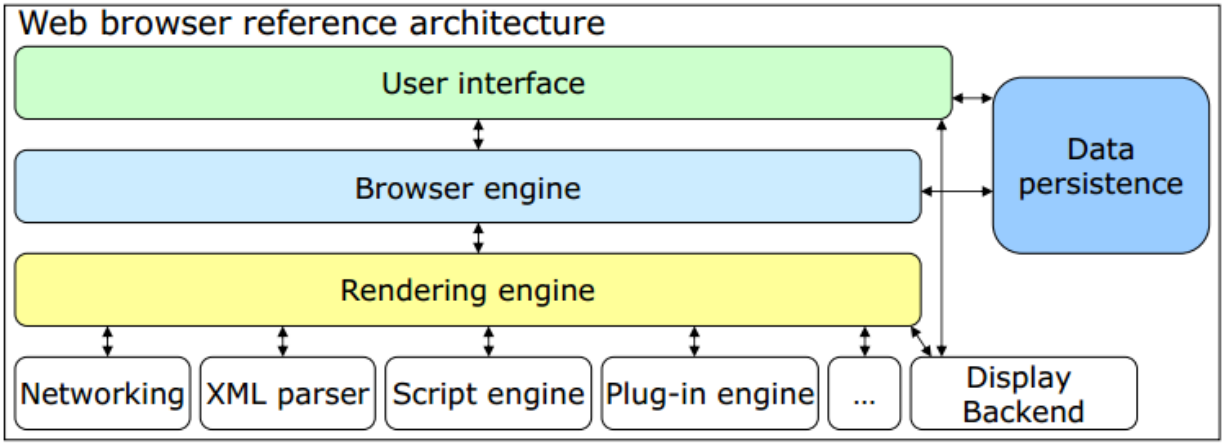


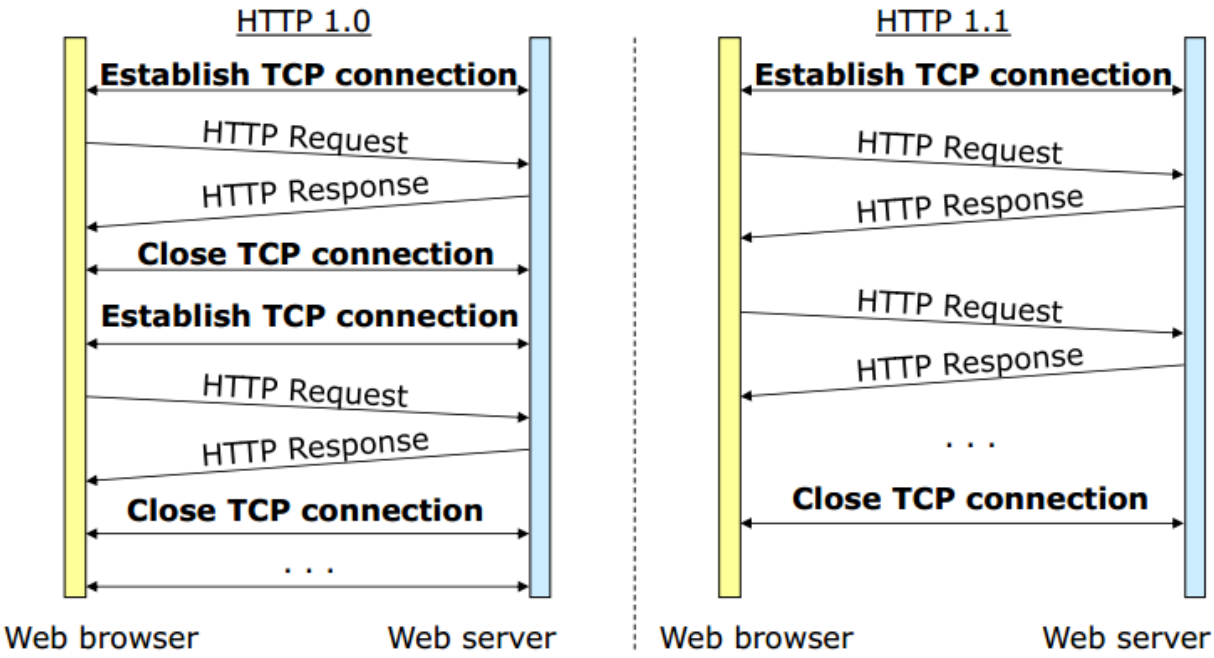
Do you remember that one? ->



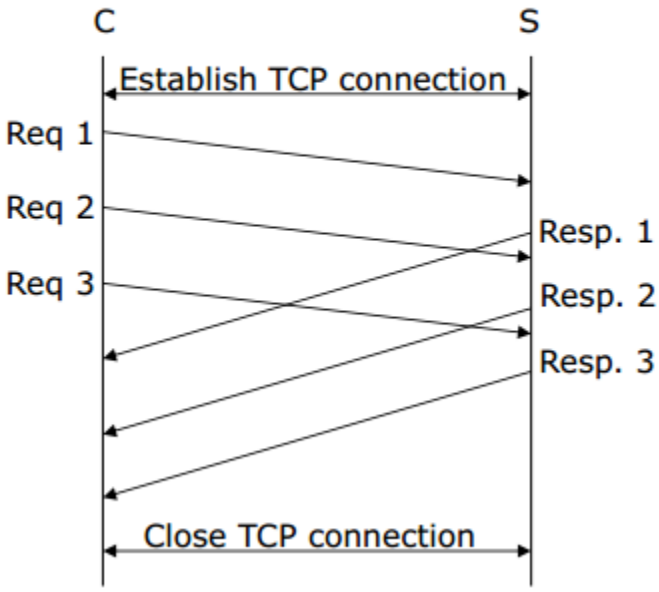
Server-side logic



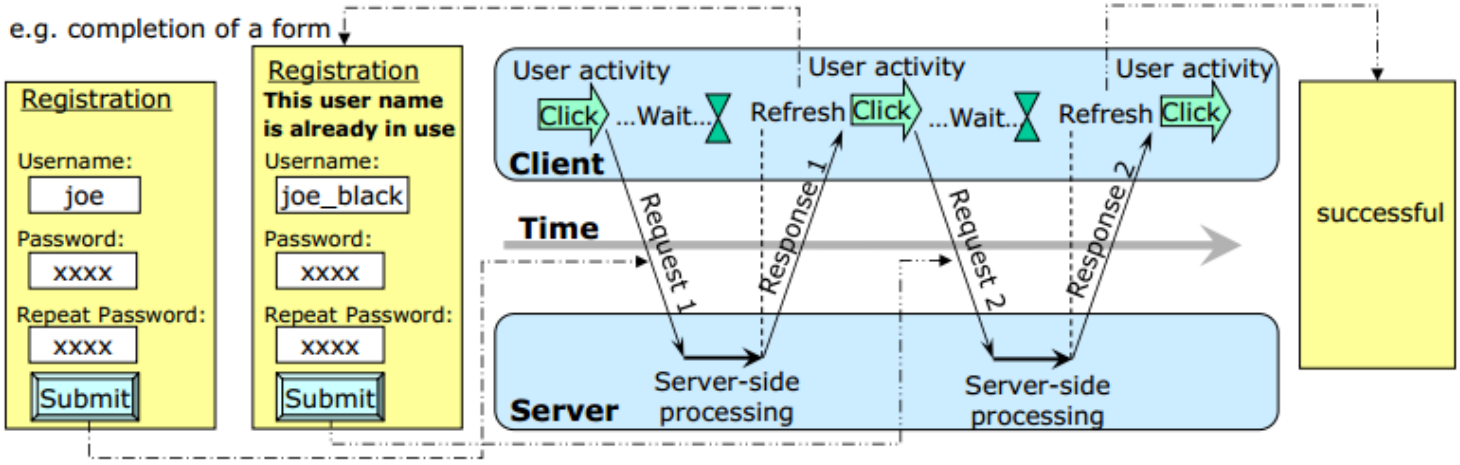




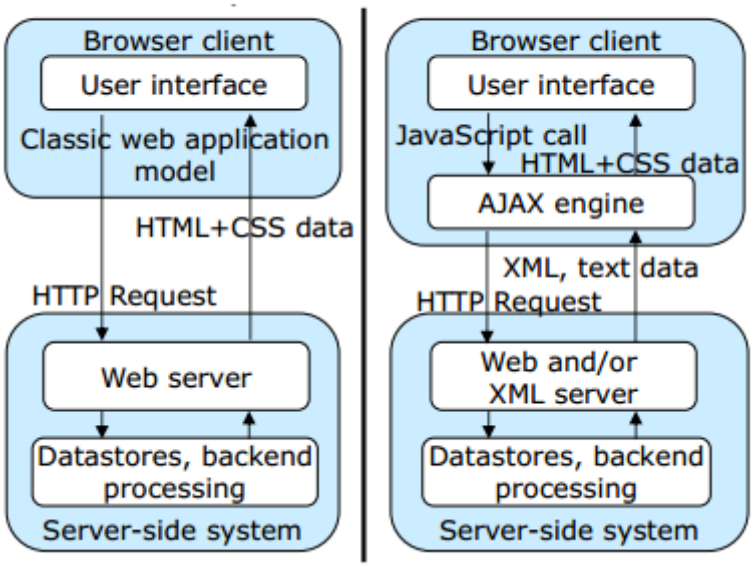
HTTP 1.1 pipelining



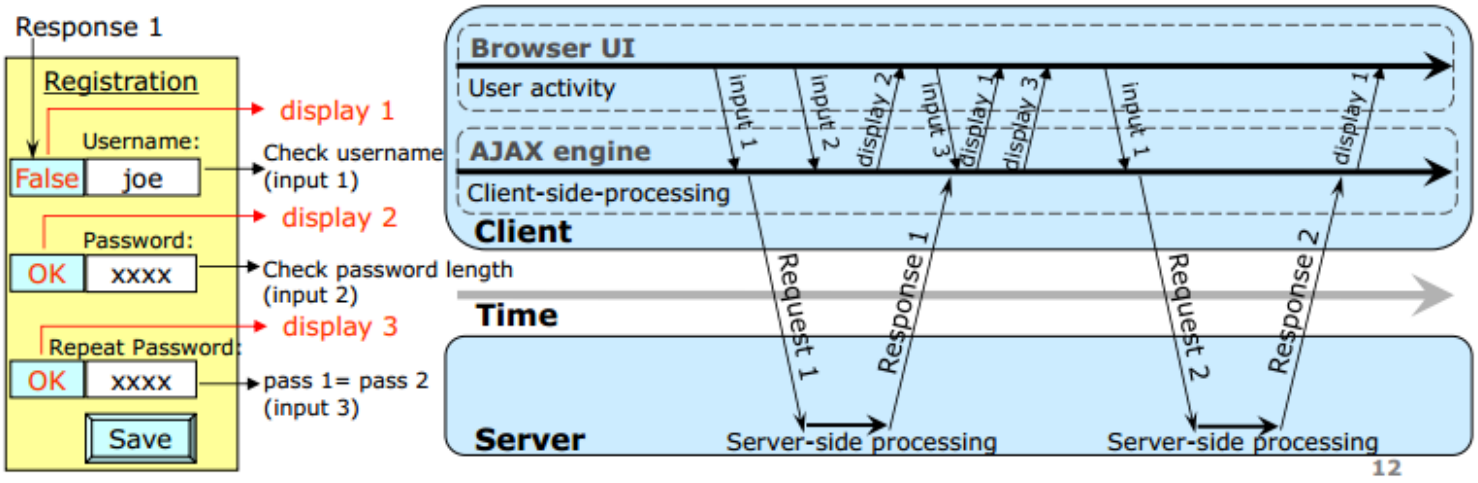
Synchronous communication model



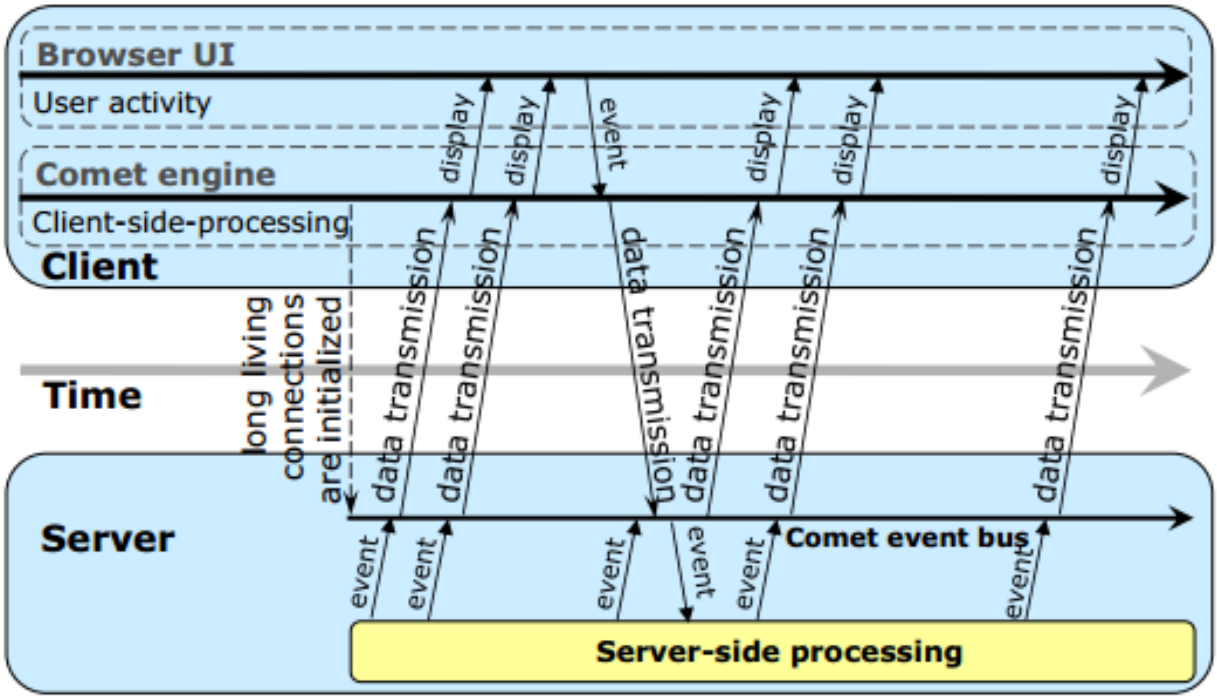
AJAX

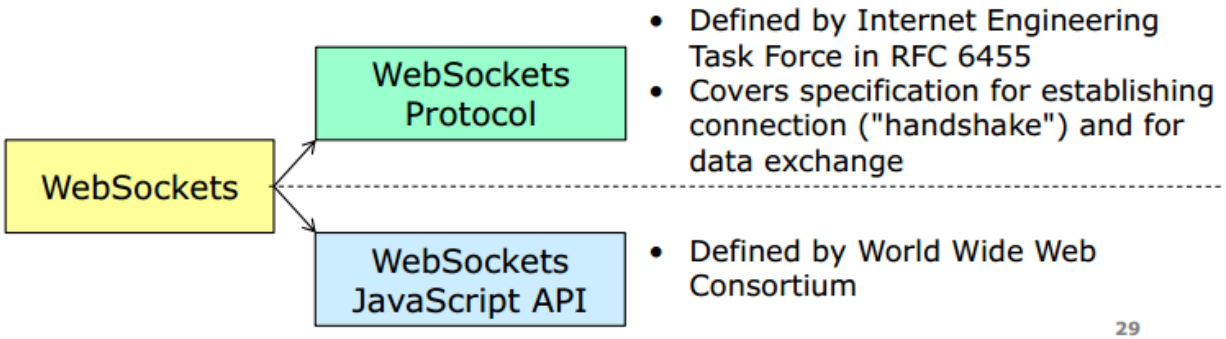


Non blocking communication model

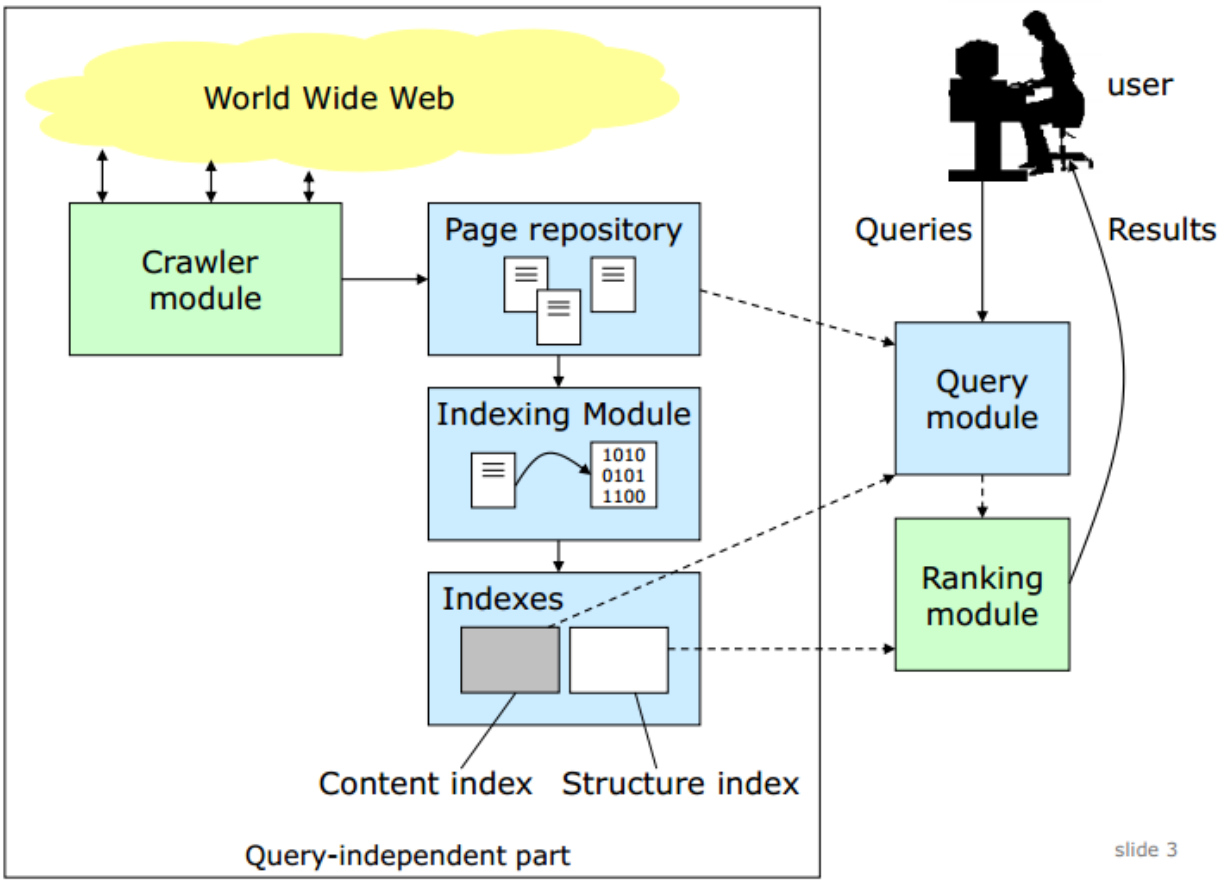


Comet – asynchronous model

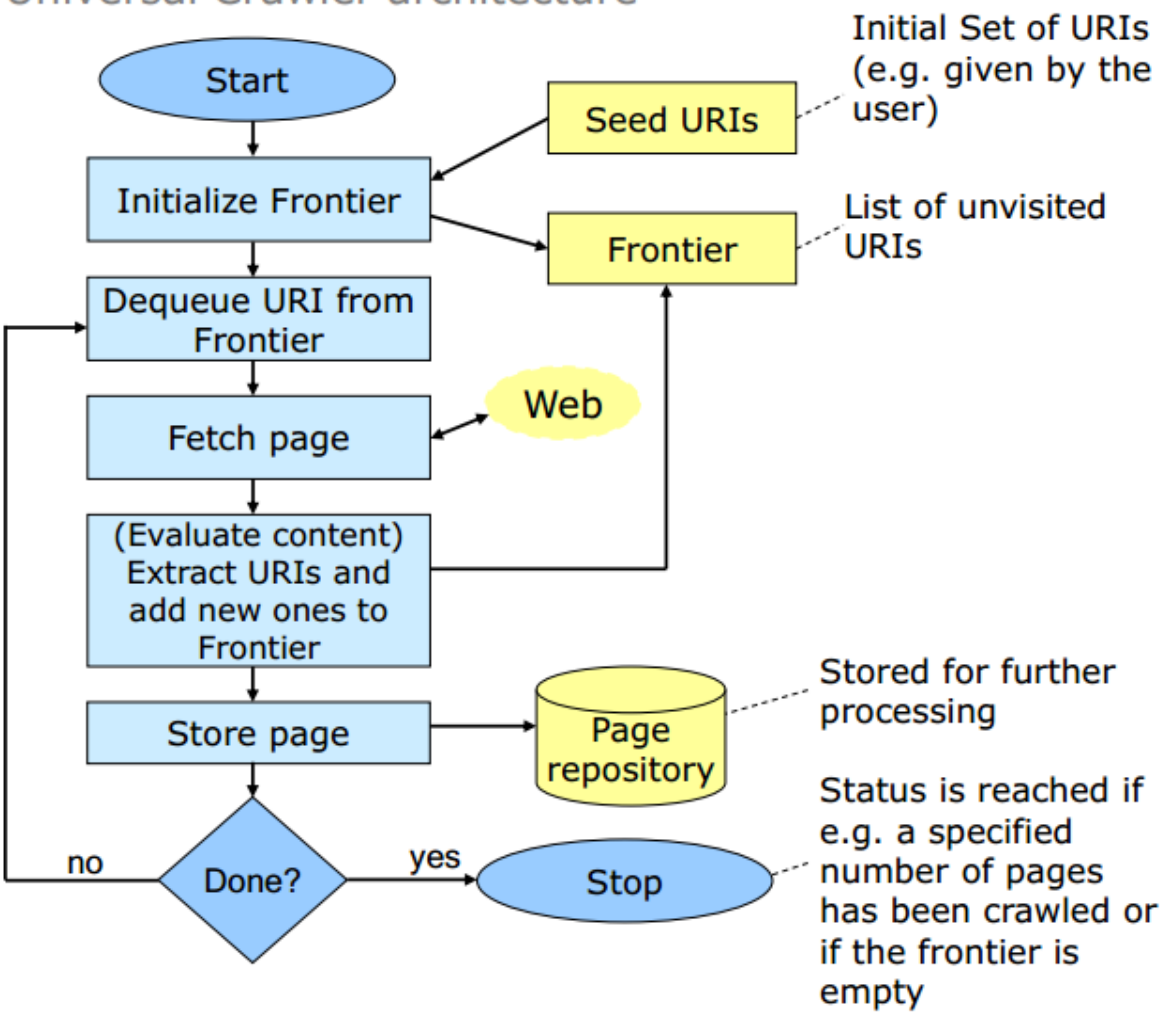


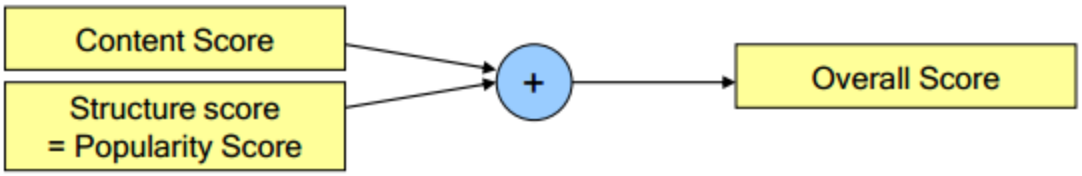


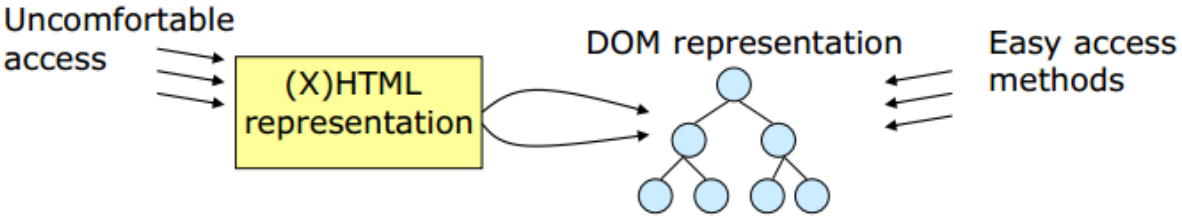
General search engine architecture



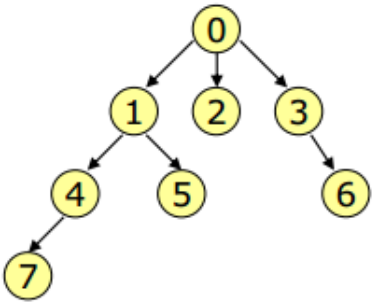
Universal Crawler architecture



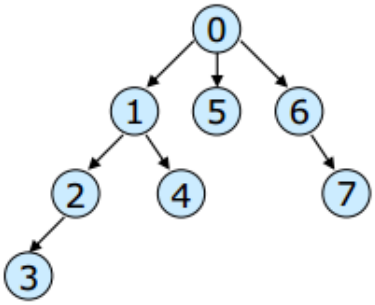


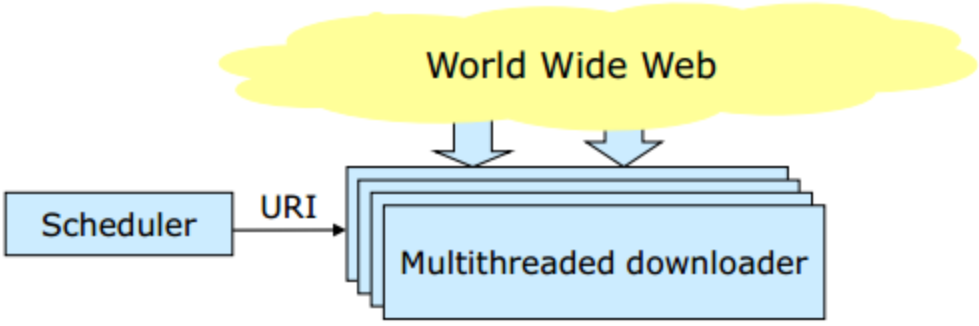


Breadth-first



Depth-first





What are you still doing here???

Isn’t that enough?!!? :D

OK, remember those:

1- XML: how to? Rules and basics.
2- SAX: how to? Rules and basics.
3- JSON: how to? Rules and basics.
4- URI/URL/URN differences and plapla.
5- HTML5 things.
6- Link extraction points.
7- adjustment factor 5-27. Is it with us? Idk..
8- Good Luck!