Information Technology

## Question 3

The internet is fraught with stuff created and posted by humans. Bits of data are uploaded and downloaded, searched, and tracked constantly across a series of computers. This search for data consists of several algorithms created and used by humans for computers. But the language of the internet is also made for humans. HTML is used to display things about a website, but it is without structure or easy readability. Linked data attempts to bring about structure to the internet by, instead of having a web of documents, it has a web of data.

Linked data is an initiative to structure data in a way such that it becomes interlinked with other data using standardised queries. Although there are already technologies designed to do this, these are built around human readability. Linked data is a technology that plans to design the internet in a way such that data can be read by computers too. This would allow data from multiple different sources connected.

Having the ability to connect so much data would be tremendously helpful when making multiple different queries. It would allow data with different formatting and sources to be easily queried and updated, which in turn would make browsing through the internet much more efficient. It would also allow for more people to use different pieces of data. For example, if a table had a person’s name and their date of birth, the information in the table would not be of any use to the reader, if they don’t already know who the person is. Let’s call the person in the table “Bea”. Bea’s name and her date of birth were written down by the creator of the table, but no other information about her can be queried. So, anyone else looking to use the table does not have enough information about it to use it properly. But if the name field (the field that contains the name “Bea”) has a link pointing to who the person is, when any future reader can then know who she is. Similarly, if the date-of-birth field has a simple string of text, such as “09-05-2003”, then readers would not be able to know if the table is using the MM-DD-YYYY format, popular in America, or the DD-MM-YYY format, popular in Britain. But if the table had an additional link, titled “birth\_date”, then readers could use this link to figure out the proper date format to use.

However, there are a few disadvantages to linked data as well. One of these is that it would be rather tedious due to the required use of publicly available links for data, instead of common words. Plus, linked data is still a new concept. Not many developers are used to it yet, and the languages involved can be difficult to get used to. However, thanks to the loss of ambiguity, additional data, and ease of use of linked data were it to be properly implemented, the benefits far outweigh the costs.

Linked data works largely on semantics. That is, the data used under this approach needs to have a valid meaning for it to work. Being able to share data easily means that this meaning can also easily be shared.

The semantic web is, in essence, the web of data previously mentioned. This would consist of a global database of “things”, where both humans and machines can easily access the data it provides. It would use technologies such as Resource Description Framework (RDF) and Web Ontology Language (OWL) to properly form the semantics used within the linked data implemented within it and enhance the usability of the internet and its services.

Several standards apply to the semantic web. For instance, the RDF is a framework that allows people to describe logical and physical objects and concepts in a way that can be processed by a computer. A PDF file is a simple document that can describe these things too, but only humans can read them. RDF files would solve this problem.

RDF files store data in “triples”, which are three-place relationships that are shaped as “subject, predicate, object”. This allows data to be stored rather easily, especially given that each part of the RDF contains a Uniform Resource Identifier (URI) link that leads to its meaning.

URIs are sequences of characters that identify a resource, abstract or physical, that is then used by the web. Entities identified by URIs can be dereferenced – that is, parts of the URI string can be broken down – and looked up by their HTTP protocol. Essentially, URIs are links to data available on the semantic web.

Another such standard is OWL. An ontology is the concept of an object that seeks its classification and explanation. OWL is a formalized vocabulary of terms that help specify the definition of terms. This is more expressive than RDF and allows for a deeper explanation for terms and relationships between terms that RDF may not be capable of.

Overall, the semantic web has tremendous potential to bring aid to the internet. Being able to access data with ease and connect to other data would bring about a much-needed change from the current web. With the standards and languages currently being implemented, there is no doubt that when Web 3.0 finally comes into fruition, the internet will be a much better place.