This is a test document to test tika-app-0.7.jar to show that it does not parse .docx Smart Tags as expected. When a .docx document contains Smart tags, several words in and around the Smart Tag do not appear in an output text document. The command run would be something like:

Java –jar /path/to/tika-app-0.7.jar –t /path/to/docx/document.docx > /path/to/output.txt.

Smart tags typically are names like George Washington, Marilyn Monroe, Napoleon Bonaparte, etc. But they are automatically generated by Word, so it can be difficult to control how they are produced. India, Thailand, Europe, Estados Unidos, Bolivia, a bunch of countries, Karl Rove, how does Word recognize a name? George Bush, Elvis Presley, Benjamin Franklin, the Eiffel Tower, Hillary Clinton, and more countries, Australia, Austria, Germany, France, Africa, New Zealand, Mexico, Italy. Yes, there are some continents in there as well, just to throw you off. Member of the WTO, IMF, and also RSPCA. And finally, another list of name that will hopefully product some more smart tags: Grant Hill, Bill Paxton, George Foreman, David Smith, Alice Hooper, and Wendy McIntyre. Here is an address (fake of course): 15 Smith St, Willhaven, AR 44512.

Now a date with time, 2pm, 15th June 2010. And some financial symbol, whatever that means, $15 USD. This is a phone number (3)15-999-0000.