Handling of Unicode

This package presents all text strings as Python unicode objects. From Excel 97 onwards, text in Excel spreadsheets has been stored as UTF-16LE (a 16-bit Unicode Transformation Format). Older files (Excel 95 and earlier) don't keep strings in Unicode; a `CODEPAGE` record provides a codepage number (for example, 1252) which is used by xlrd to derive the encoding (for same example: "cp1252") which is used to translate to Unicode.

If the `CODEPAGE` record is missing (possible if the file was created by third-party software), `xlrd` will assume that the encoding is ascii, and keep going. If the actual encoding is not ascii, a `UnicodeDecodeError` exception will be raised and you will need to determine the encoding yourself, and tell xlrd:

```
book = xlrd.open_workbook(..., encoding_override="cp1252")
```

If the `CODEPAGE` record exists but is wrong (for example, the codepage number is 1251, but the strings are actually encoded in koi8_r), it can be overridden using the same mechanism.

The supplied `runxlrd.py` has a corresponding command-line argument, which may be used for experimentation:

```
runxlrd.py -e koi8_r 3rows myfile.xls
```

The first place to look for an encoding, the "codec name", is the Python documentation.