Report on the python hashing implementation.

General Overview:

pyHasher normalizes a file and creates a hash for it. The normalization is done by running mnfy3 or mnfy2, and if both fail, falling back to mnfyfailsafe, then creating a sha512 hash of the output file. It can be run in either python2 or python3, as long as both are installed.

Usage instructions:

To hash a single file, use the following command:

‘pyHasher.py <filename>’

Optionally, to see the normalized file, append ‘—dump’. (though this is only useful for debugging/test)

Requirements/Install:

The code is self contained, and uses relative paths to access everything. However the entry point function pyHasher assumes that both Python 2.7 and Python 3.3 are installed. The python environment names need to be set in the config variables at the top of pyHasher.py.

Assumptions:

Differences between ASTs:

Python 2.4	Python 2.5	Python 2.6	Python 3.3
*Cant be found/Does not exist	ImportFrom module was not optional in 2.5 Set/SetComp/DictComp were introduced in 2.7	ImportFrom module was not optional in 2.5 Set/SetComp/DictComp were introduced in 2.7	Guide: + implies only in 3.3 - implies only in 2.7 ~ implies in both 2.7 and 3.3 but different. +WithItem +Starred +Bytes +YieldFrom +NonLocal +Arg +Annotations ~ClassDef ~FunctionDef ~Raise ~ExceptHandler ~With ~Arguments ~TryExcept/TryFinally/Try

			-Print -Exec -Repr

Changes Made and Status:

Change	Developed	Tested	Notes
WithItem	N/A	N/A	Removed
Starred	N/A	N/A	Removed
Bytes	N/A	N/A	Removed
YieldFrom	N/A	N/A	Removed
NonLocal	N/A	N/A	Removed
Arg	N/A	N/A	Removed
Annotations	N/A	N/A	Removed
ClassDef	✓	✓	Keywords, *args and kwargs Removed.
Function Def	✓	✓	Returns removed
Raise	✓	✓	Tracebacks added
ExceptHandler	N/A	N/A	Both expr? And identifier? Handled in same way. No Change required.
With	✓	✓	WithItem replaced. However the safe-transformation “CombineWithStatements” was not ported to 2.7
Arguments	✓	✓	Keyword only args, annotations and default keyword args removed.
Try [Except/Finally]	✓	✓	Try split into Try Except and Try Finally. However the safe-transformation “visit_TryExcept” in “EliminateUnusedConstants” has not been ported.
Print	✓	✓	Works fine except: Print(‘hello’) will become print ‘hello’ in mnfy2.7 But remain print(‘hello’) in mnfy3.3.
Exec	✓	✓	Added
Repr	✓	N/A	Repr is theoretically implemented, but ast.parse() treats repr as a function call instead of a node, so its not possible to test. (The code is never executed)

Known Issues:

#	Issue	Location	Impact	Severity	Action
1	Slicing error- P=(p+[1,2,3])[:1] Gets minified to P=p+[1,2,3][:1] Which has different meaning.	Mnfy27, Mnfy33	No impact on hashing because its consistant. Is an issue for mnfy's standalone goals of portable code.	Low	No action required for hashing /RedHat purposes.
2	Some docstrings are referenced by __doc__ call in python. Docstrings are removed by mnfy, causing compile/run problems.	Mnfy27, Mnfy33	No impact on hashing because code does not need to compile and removal of docstrings is consistent.	Low	No action required for hashing/RedHat purposes.
3	Big Numbers: For some reason, reading and printing very large numbers (in the order of 10**10) causes digits to be lost.	Mnfy27, but I think it is in all of Python.	Potential hashing risk, because large numbers could change code execution, though both hashes would be the same.	Medium	Not really sure if there is a fix for this one.

Testing Report:

Test scripts:

/libs/mnfy33/test_mnfy.py - Unit Testing of Mnfy33 (written by Brett Cannon)

/libs/mnfy33/verify_mnfy.py - Coverage Testing (written by Brett Cannon)

/libs/mnfy273/test_mnfy.py - Unit Testing of Mnfy27 (modified from mnfy3300 version)

- Fails on test_empty_Try because unused try transform has not been ported.

/libs/mnfy273/verify_mnfy.py - Coverage Testing of Mnfy27 (modified from mnfy3300 version)

- Fails on aifc.py because of Issue#3 and on urllib because of Issue#1.

/libs/mnfyFailSafe/test_mnfy.py - Unit Testing of MnfyFailSafe

testHasher.py – Testing of hashing function/blackbox testing/testing multiple environments.

*Note: Not all tests pass, because of the issues listed above.