

Table of Contents

1 Overview	1
2 Deliverables/Tasks.....	2
3 Detailed Description	2
3.1 Conditions to check for any input type	3
3.1.1 Filename checks	5
3.1.2 Metadata checks	5
3.1.3 If the document is an RFC.....	6
3.1.4 If the document is an Internet-Draft (that is, not an RFC)	6
3.2 XML Input Specific Conditions	6
3.2.1 Boilerplate checks.....	7
3.3 Text Input Specific Conditions.....	7
3.3.1 Boilerplate checks.....	7
3.3.2 If the document is an Internet-Draft (i.e not an RFC)	8
4 Functionality from the current implementation that will not be carried forward	9

1 Overview

The idnits program inspects Internet-Draft documents for a variety of conditions that should be adjusted to bring the document into line with policies from the IETF, the IETF Trust, and the RFC Editor. Several large upcoming changes in those policies need to be reflected. In particular, the RFC Editor will allow non-ASCII UTF-8 to appear in current text format RFCs, and the current program treats the presence of such characters as an error. As RFCs transition to publication in the xml2rfc v3 format, the xml source for those RFCs (and Internet-Drafts that lead to them) should be inspected directly.

The current program has grown through incremental improvement over several years. Some very complex logic has been necessary at times, such as during the transition to the current Trust Legal Provisions (TLP) policies. It is written as a shell script that invokes several utilities, primarily awk. Rather than continuing to expand this script, the current maintainer recommends a fresh implementation, in Python, with the following high-level goals:

- Modularity, facilitating integration with the datatracker for document metadata, and draft submission
- Simplicity, removing much of the accumulated complexity
- Maintainability, drawing on the skills of the growing set of volunteers working with the datatracker codebase

2 Deliverables/Tasks

- Identification of high-level modules and design of the API for those modules
- Identification appropriate existing APIs into document and series metadata (for retrieving current document state and the contents of the downref registry), along with design and development of any additional needed interfaces.
- Design of the command-line arguments
- Development of the application
- Development of an extensible suite of test documents and tests demonstrating correct behavior for each of the below requirements
- Integration of the application into production systems

3 Detailed Description

The program will accept either a text document or an xml document that uses the xml2rfc v3 grammar. It will produce diagnostic output on stdout appropriate for direct display, and for inclusion in a web page.

The program will be configurable to operate in several modes, affecting how severely to consider each issue found. The appropriate severity for each issue is detailed later in this document.

- Normal: Apply the full set of checks with the default severity
- Forgive Checklist Issues: Treat checklist issues as less severe
- Submission: Only treat issues that should prevent submission as errors

The program will have options allowing:

- Different levels of output verbosity, including a debug level that will present diagnostics of internal operation (such as displaying text matches)
- Suppression of issues below a given level of severity (e.g. suppress warnings and comments)

The program will be resilient to files created on various operating systems with different conventions for line-endings. (It will treat CR, LF, and CRLF as equivalent.)

The program must be usable on modern UNIX-like systems (including Linux and OS/X), and modern versions of Microsoft Windows. The program is expected to use the Internet to retrieve information from the datatracker and possibly other systems, but must remain usable when the host has no Internet connection. If information needed for diagnostics (such as the current status of a referenced RFC) is not available, the program will note that in its output and proceed without it.

The program will be used as part of a web service, replacing the service currently available at <http://www.ietf.org/tools/idnits>.

If the program is presented with an RFC and asked to operate in the submission mode, it will produce an error and stop processing.

If the program is presented with a text input document that appears to be a tombstone, it will report so and stop processing.

If the program is presented with an XML input document that is not well formed, it will process as much as possible, and provide diagnostics on where the format becomes invalid.

When producing verbose output, the program will list each occurrence of an issue, along with where the issue occurs. For all input document types, it will note the line number of the line exhibiting the issues. For xml input documents, it should, when possible, also provide an XPath expression that selects the content with the issue as specifically as possible.

To test for many of the conditions described later in this document, the program will need to identify a section or attribute from the input document. If the input document is a text document, the program is expected to extract the needed information in ways informed by the current program (the algorithms there have been tuned over many years). When the input is an xml document, the needed information will often be directly identified. For example, the abstract will be within the <abstract> tag in the xml document. Other section headings can be searched for in a <titleelement> tag with a <section> tag. The document's date will be in a <date> child of <front>.

Several of the conditions below speak of “sufficiently matching” given text. The program will follow the algorithms in the current implementation to make that determination.

In the below tables, the three modes and the severity levels have been abbreviated to improve the layout of the condition descriptions. The abbreviations are as follows

- norm: normal
- f-c: forgive-checklist
- sub: submission
- warn: warning
- comm: comment

If a row ends with a “*”, the treatment of the condition is either new, or is significantly different from the current idnits policies.

3.1 Conditions to check for any input type

Condition	norm	f-c	sub	
Control characters other than CR, NL, or FF appear (<i>0x01-0x09,0x0b,0x0e-0x1f</i>)	error	warn	none	
Byte sequences that are not valid UTF-8 appear	error	warn	none	*
Non-ASCII UTF-8 appear (<i>comment will point to guidance in draft-flanagan-nonascii</i>)	comm	comm	none	*

Missing Abstract section	error	error	error
Missing Introduction section	error	warn	none
Missing Security Considerations section	error	warn	none
Missing Author Address section	error	warn	none
References (if any present) are not categorized as Normative or Informative	error	warn	none
Abstract contains references	error	warn	none
FQDN appears (other than www.ietf.org) not meeting RFC2606/RFC6761 recommendations	warn	warn	none
Private IPv4 address appears that doesn't meet RFC5735 recommendations	warn	warn	none
Multicast IPv4 address appears that doesn't meet RFC5771/RFC6676 recommendations	warn	warn	none
Other IPv4 address appears that doesn't meet RFC5735 recommendations	warn	warn	none
Unique Local IPv6 address appears that doesn't meet RFC3849/RFC4291 recommendations	warn	warn	none
Link Local IPv6 address appears that doesn't meet RFC3849/RFC4291 recommendations	warn	warn	none
Other IPv6 address appears that doesn't meet RFC3849/RFC4291 recommendations	warn	warn	none
A possible code comment is detected outside of a marked code block	warn	warn	warn
2119 keywords occur, but neither the matching boilerplate nor a reference to 2119 is missing	error	warn	none
2119 keywords occur, a reference to 2119 exists, but matching boilerplate is missing	warn	warn	none
2119 boilerplate is present, but document doesn't use 2119 keywords	warn	warn	none
badly formed combination of 2119 words occurs (<i>MUST not, SHALL not, SHOULD not, not RECOMMENDED, MAY NOT, NOT REQUIRED, NOT OPTIONAL</i>)	comm	comm	none
text similar to 2119 boilerplate occurs, but doesn't reference 2119	error	error	none

NOT RECOMMENDED appears, but is not included in 2119-like boilerplate	warn	warn	none	
Abstract doesn't directly state it updates or obsoletes each document so affected (<i>Additional comment if Abstract mentions the document some other way</i>)	comm	comm	none	
Abstract states it updates or obsoletes a document not declared in the relevant field previously	comm	comm	none	
Author's address section title misuses possessive mark or uses a character other than a single quote	warn	warn	none	
a reference is declared, but not used in the document	warn	warn	warn	
a reference appears to be a downref (noting if reference appears in the downref registry)	error	warn	none	
a normative reference to an document of unknown status appears (<i>possible downref</i>)	comm	comm	none	
a normative or unclassified reference is to an obsolete document	error	warn	none	
an informative reference is to an obsolete document	comm	comm	none	
a reference is to a draft that has already been published as an rfc	warn	warn	none	
A code-block is detected, but the block does not contain a license declaration	warn	warn	none	*
3.1.1 Filename checks				
filename's base name contains characters other than digits, lowercase alpha, and dash	error	error	error	*
filename's extension doesn't match format type (.txt, .xml)	error	error	error	*
filename's base name doesn't match the name declared in the document	error	error	error	*
filename (including extension) is more than 50 characters	error	error	error	
3.1.2 Metadata checks				
Document claims to obsolete an RFC that is already obsolete	warn	warn	none	*
Document claims to update and RFC that is obsolete	warn	warn	none	*
Document's status or intended status is not found or not recognized	warn	warn	warn	

Document's date can't be determined or is too far in the past or the future (<i>see existing implementation for "too far"</i>)	warn	warn	warn	
--	------	------	------	--

3.1.3 If the document is an RFC

Missing IANA considerations section	comm	comm	N/A	
-------------------------------------	------	------	-----	--

3.1.4 If the document is an Internet-Draft (that is, not an RFC)

filename's base name doesn't begin with 'draft', contains two consecutive hyphens, or doesn't have enough structure to contain the individual or stream, potentially a wg name, and a distinguishing name. (<i>draft-example-00 is an error, but draft-example-filename is acceptable</i>)	error	error	error	
--	-------	-------	-------	--

Missing IANA considerations section	error	warn	none	
-------------------------------------	-------	------	------	--

3.1.4.1 Additional metadata check

version of document is unexpected (already exists, or leaves a gap)	warn	warn	warn	*
---	------	------	------	---

3.2 XML Input Specific Conditions

any deprecated elements or attributes appear	warn	warn	warn	*
--	------	------	------	---

metadata and document's 'submissionType' attribute state different streams	error	error	error	*
--	-------	-------	-------	---

The text inside a <sourcecode> tag contains the string '<CODE BEGINS>' (<i>Warn that the string is unnecessary and may duplicate what a presentation format converter will produce.</i>)	warn	warn	none	*
--	------	------	------	---

The text inside any other tag contains the string '<CODE BEGINS>' (<i>Warn that if the text is a code block, it should appear in a <sourcecode> element</i>)	warn	warn	none	*
--	------	------	------	---

text occurs that looks like a text-document reference (e.g. [1], or [RFC...]) (<i>if the text was really a reference it should be in an <eref> tag</i>)	warn	warn	none	*
---	------	------	------	---

<rfc> ipr attribute is missing or not recognized	error	error	error	*
--	-------	-------	-------	---

ipr attribute is not one of "trust200902", "noModificationTrust200902", "noDerivativesTrust200902", or "pre5378Trust200902"	warn	warn	warn	*
---	------	------	------	---

document is ietf stream and ipr attribute is one of “noModificationTrust200902” or “noDerivativesTrust200902”	error	error	error	*
---	-------	-------	-------	---

3.2.1 Boilerplate checks

The text inside any tag sufficiently matches any of the boilerplate in the IETF-TLP-4 section 6a-6d (<i>such text should probably be removed and the ipr attribute of the rfc tag should be verified</i>)	warn	warn	warn	*
---	------	------	------	---

The value of the <rfc> autogeneratedBoilerplateText, if non-empty, does not match what the ipr attribute would cause to be generated	warn	warn	warn	*
--	------	------	------	---

3.3 Text Input Specific Conditions

document does not appear to be ragged-right (<i>more than 50 lines of intra-line extra spacing</i>)	error	warn	none	
---	-------	------	------	--

document contains over-long lines (<i>cut-off is 72 characters. Report longest line, and count of long lines</i>)	error	warn	warn	
---	-------	------	------	--

document has hyphenated line-breaks	warn	warn	none	
-------------------------------------	------	------	------	--

Updates or Obsoletes line on first page has more than just numbers of RFCs (such as the the character sequence ‘RFC’)	warn	warn	none	
---	------	------	------	--

Document starts with PK or BM	comm	comm	none	
-------------------------------	------	------	------	--

document appears to use numeric references, but contains something that looks like a text-style reference (or vice-versa)	comm	comm	none	
---	------	------	------	--

a string that looks like a reference appears but does not occur in any reference section	warn	warn	warn	
--	------	------	------	--

Abstract section is numbered	error	error	error	
------------------------------	-------	-------	-------	--

‘Status of this memo’ section is numbered	error	error	error	
---	-------	-------	-------	--

Copyright Notice section is numbered	error	error	error	
--------------------------------------	-------	-------	-------	--

3.3.1 Boilerplate checks

TLP-4 6.b.i copyright line is not present	error	error	error	
---	-------	-------	-------	--

TLP-4 6.b.i copyright date is not this (or command-line specified) year	warn	warn	warn	
---	------	------	------	--

TLP-4 6.b.i or b.ii license notice is not present, or doesn’t match stream	error	error	error	
--	-------	-------	-------	--

IETF stream document sufficiently matches TLP-4 6.c.i or 6.c.ii text (restrictions on publication or derivative works)	error	error	error
More than one instance of text sufficiently matching the TLP-4 6.b.i copyright line occurs	warn	warn	warn
More than one instance of text sufficiently matching either the TLP4 6.b.i or 6.b.ii license notice occurs	warn	warn	warn
Document obsoletes or updates any pre-5378 document, and doesn't contain the pre-5378 material of TLP4 6.c.iii	warn	warn	warn
Any prior version of the document might be pre-5378 and the document doesn't contain the pre-5378 material of TLP4 6.c.iii	warn	warn	warn
3.3.2 If the document is an Internet-Draft (i.e not an RFC)			
contains over-long pages	warn	warn	none
report count of pages with more than 58 lines			
doesn't say INTERNET DRAFT in the upper left of the first page	error	error	error
doesn't have expiration date on first and last page	error	error	error
doesn't have an acceptable paragraph noting that IDs are working documents	error	error	error
doesn't have an acceptable paragraph calling out 6 month validity	error	error	error
doesn't have an acceptable paragraph pointing to the list of current ids	error	error	error
has multiple occurrences of current id text	error	error	error
document name doesn't appear on first page	error	error	error
has more than 15 pages and not Table of Contents	error	error	warn
IPR disclosure text (TLP 4.0 6.a) does not appear	error	error	error
IPR disclosure text (TLP 4.0 6.a) appears after first page	error	error	error
pages are not separated by formfeeds	warn	warn	none
'FORMFEED' and '[Page' occur on a line, possibly separated by spaces (<i>indicates NROFF post-processing wasn't successful</i>)	comm	comm	comm
section title occurs at an unexpected indentation	warn	warn	none

4 Functionality from the current implementation that will not be carried forward

The current implementation has several undocumented command-line options that will not be carried forward. In particular, the new implementation will not perform spell checking or grammar checking. When assessing text input, the new implementation will not check for a Shadow-directories line.