
JavaScriptin staattinen tyypittäminen

LuK -tutkielma
Turun yliopisto
Tulevaisuuden teknologioiden laitos
Tietojenkäsittelytiede
2017
Oskari Noppa

Sisältö

1	Johdanto	1
2	Peruskäsitteitä	3
2.1	Tyyppijärjestelmien luokittelu	3
2.2	ECMA-262, EcmaScript ja JavaScript	4
2.3	TypeScript	5
2.4	Flow	5
2.5	Closure kääntäjä	6
3	JavaScript-ohjelman muuttaminen staattisesti tyyppitarkastetuksi	7
3.1	Tyyppiannotaatiot	7
3.2	Ohjelman kääntäminen ennen suorittamista	9
3.3	Työkalun vaiheittainen käyttöönotto	11
4	Virheiden havaitseminen	13
4.1	Virheiden havaitseminen	13
4.2	Ohjelman optimointi käännösvaiheessa	14
4.3	Tyypin määritys dokumentaationa	14
5	Ongelmat JavaScriptin staattisessa tyyppittämisessä	15
5.1	Vaikeasti analysoitavat suunnittelumallit	15
5.2	Tyypiturvallisuus ja käytännöllisyys	15

6 Yhteenveto	16
Lähdeluettelo	17

Luku 1

Johdanto

Kaikenlaisen ohjelmoinnin keskiössä on data ja yksi tärkeimmistä datan ominaisuuksista on sen tyyppi. Muuttuja “nimi” voi olla datatypiltään teksti ja muuttuja “ikä” voi olla numero, eikä näitä kahta voi huolettomasti sekoittaa. Ohjelman tila ei ole järkevä jos se sanoo henkilön iän olevan “Matti”. Ohjelmointikielet voidaan hyvin karkeasti jakaa kahteen sen mukaan miten ja missä vaiheessa niissä käsitellään muuttujien tyyppejä.

Staattisesti tyypitetyiksi kutsutaan kieliä jotka vaativat että ohjelman käsittelemien tietorakenteiden tyypit on tulkittavissa käännös aikana, eli jo ennen ohjelman suorittamista. Lähes aina ohjelmointikieli saa tämän tiedon tietotyypeistä vaatimalla koodiin erityisiä tyypinmäärittelyitä. Ohjelman koodissa voitaisiin esimerkiksi määrittää että henkilön ikä on aina numero. Tällöin ohjelmointikielen kääntäjään rakennetut tarkistukset voivat vah-
tia jo ennen koodin suorittamista ettei ohjelmoija virheellisesti yritä asettaa henkilön iäksi tekstiä tai mitään muutakaan ei-numeerista arvoa. Dynaamisesti tyypitetyissä kielissä sen sijaan vastuu oikeiden tietotyyppien käyttämisestä jätetään ohjelmoijalle, kieli ei vaa-
di kehittäjältä tyyppien eksplisiittistä määrittämistä eikä niiden oikeellisuutta tarkasteta ainakaan ennen ohjelman suorittamista.

Kyseessä on kielen suunnittelullinen valinta. Kielen dynaaminen tyypittäminen vaa-
tii yleensä vähemmän varsinaisen koodin kirjoittamista saman tuloksen saavuttamiseksi, sillä erillisiä tyypinmäärittelyitä ei tarvitse kirjoittaa. Toisaalta staattinen tyypitys mah-

dollistaa monia hyödyllisiä kehitystyökaluja ja auttaa poistamaan väärin tietotyyppien käyttämisestä johtuvan bugien osajoukon. Kielen suunnittelijan on löydettävä haluamansa tasapaino kielen ominaisuuksien välillä ja arvioitava mikä on parhaaksi niihin käyttö-tarkoituksiin joihin kielen on tarkoitus hyvin soveltua.

Luku 2

Peruskäsitteitä

2.1 Tyyppijärjestelmien luokitteleminen

Ohjelmointikielten tyyppijärjestelmien jakaminen staattisesti ja dynaamisesti tyypitarkastettuihin perustuu ohjelman kehitysvaiheeseen jossa tarkastaminen tapahtuu. Staattisella tyypitarkastamisella viitataan ohjelman tyyppien analyysiin ennen sen suorittamista, esimerkiksi käännösaikana, kun taas dynaaminen tyypitarkastus varmistaa arvojen tyyppien oikeellisuuden ohjelmaa suoritettaessa. Tyyppijärjestelmät voidaan jaotella myös muiden ominaisuuksien perusteella, esimerkiksi vahvoihin ja heikkoihin tyyppijärjestelmiin. Näiden termien merkitys ei ole tarkasti määritelty, mutta yleisesti niillä viitataan tapaan jolla kieli käsittelee tarkoitettua poikkeavat, virheelliset tyypit [1]. Vahvasti tyypitetssä kielessä tällainen aiheuttaisi käännös- tai ajonaikaisen virheen, kun taas heikosti tyypitetssä kielessä arvoille voitaisiin tehdä implisiittisiä tyypimuunnoksia niiden yhteensopivuuden saavuttamiseksi.

JavaScript on dynaamisesti tarkastettu, heikosti tyypitetty kieli. Esimerkiksi ohjelma `"teksti".potenssiin(3)` antaa staattisesti tyypitarkastetussa kielessä virheen jo käännösaikana, mikäli metodia `potenssiin` ei ole tekstityypisille arvoille määritetty. JavaScriptiä suorittava ympäristö sen sijaan hyväksyisi ohjelman ja sallisi sen suorittamisen. Virhe olemattoman metodin kutsumisesta ilmenisi vasta jos ohjel-



Kuva 2.1: Tyyppijärjestelmät eri ohjelmointikielissä

man suoritus evaluoi kyseisen ilmaisen. Lisäksi esimerkiksi ilmaisu `"teksti" + 2` ei aiheuttaisi virhettä edes suoritusaikana, sillä heikoille tyyppijärjestelmille ominaisesti JavaScript muuttaisi numeron 2 string-muotoon ennen summausoperaation arviointia. Tässä tutkielmassa keskitytään lähinnä JavaScriptin tyyppien staattiseen ja dynaamiseen, eli käytännössä käännös- ja ajonaikaiseen tarkastamiseen. Jotkin esitellyistä työkaluista myös tiukentavat kielen sallimia operaatioita siten, että esimerkiksi yllä esitettyä `"teksti" + 2` ohjelmaa ei enää sallittaisi. Monia muita heikoille tyyppijärjestelmille tavallisia ominaisuuksia jää kuitenkin tarkistamatta.

2.2 ECMA-262, EcmaScript ja JavaScript

“EcmaScript” on ECMA-262 standardin määrittelemä ohjelmointikieli [2, 3], jonka kehityksestä vastaa organisaatio Ecma International. “JavaScript” on Oraclen omistama tavaramerkki jolla viitataan EcmaScript-kielen osittaisiin tai täydellisiin toteutuksiin.[4] Historiallisista syistä termejä “JavaScript” ja “EcmaScript” käytetään usein keskenään vaih-

tokelpoisesti. Tässä tutkielmassa termillä “JavaScript” viitataan ECMA-262:den kahdeksannen version mukaiseen EcmaScriptiin, jota kutsutaan myös nimellä EcmaScript 2017.

2.3 TypeScript

TypeScript on Microsoftin luoma ohjelmointikieli, jonka tarkoitus on auttaa JavaScript-ohjelmien kehitystä staattisen tyyppijärjestelmän avulla. Se on EcmaScriptin ylijoukko (superset) [4] ja jatkaa JavaScriptin syntaksia tyyppimäärittelyihin käytettävällä annotaationsyntaksilla. Jokainen validi JavaScript-ohjelma on syntaksiltaan ja ajonaikaiselta käyttäytymiseltään validi TypeScript-ohjelma. TypeScript kuitenkin lisää kehitykseen käännösvaiheen, jossa ohjelman tyyppien oikeellisuus tarkastetaan staattisesti. TypeScript koodi käännetään JavaScriptiksi, joka puolestaan voidaan suorittaa selaimissa tai muissa JavaScriptin suoritussympäristöissä. TypeScript kääntäjän voi myös määrittää muokkaamaan tulostettava koodi yhteensopivaksi vanhempien EcmaScript-standardien kanssa, mikä on hyödyllistä jos ohjelman on tarkoitus tukea sellaisia suoritussympäristöjä jotka eivät tue uusinta EcmaScriptin versiota.

2.4 Flow

Flow on Facebookin kehittämä työkalu, joka TypeScriptin tavoin jatkaa JavaScriptin syntaksia staattisesti tarkastettavilla tyyppimäärittelyillä. Flow itsessään ei sisällä kääntäjää, vaan keskittyy yksinomaan ohjelman tyyppiturvallisuuden tarkastamiseen. Koodiin lisätyt tyyppimäärittelyt on kuitenkin poistettava ennen kuin JavaScript-ohjelma voidaan suorittaa. Tähän tarkoitukseen voidaan käyttää esimerkiksi Babel-kääntäjää, joka poistaa Flow-tyyppimäärittelyt ja muokkaa JavaScript-koodin yhteensopivaksi toivotun EcmaScript-version kanssa [5].

2.5 Closure kääntäjä

Googlen Closure kääntäjä on käännöstyökalu, jonka pääasiallinen tarkoitus on minimoida ja optimoida JavaScript-koodia käännösvaiheessa ennen tuotantoon siirtämistä. Closure sisältää kuitenkin myös tuen tyyppivirheiden tarkastamiselle käännösvaiheessa [6]. Tyypit annotoidaan erityisellä JSDoc-pohjaisilla dokumentaatiokommenteilla. Koska annotaatiot ovat kommenteissa eivätkä erityisenä syntaksina muun suoritettavan koodin joukossa, Closure-annotoitua JavaScriptiä ei tarvitse kääntää ennen sen suorittamista [7].

Luku 3

JavaScript-ohjelman muuttaminen staattisesti tyyppitarkastetuksi

3.1 Tyyppiannotaatiot

Tyyppijärjestelmä voi päätellä muuttujan sallitun tyylin automaattisesti päättelemällä tai kieleen sisältyvien eksplisiittisten tyyppimäärittelyjen perusteella. Kaikki kolme tässä esiteltyä JavaScriptin staattiseen tyyppitarkastukseen tarkoitettua työkalua päättelevät muuttujien tyyppiä automaattisesti, mutta vaativat myös eksplisiittisiä määrytyksiä paikoitellen. Closure-kääntäjä lukee tyyppimäärytykset JSDoc-tyylisistä dokumentaatiokommenteista [7].

```

1  /**
2   * @param {!Array<Ostos>} ostokset
3   * @return {number} Ostosten yhteenlaskettu hinta
4   */
5  function ostoskorinHinta(ostokset) {
6      let summa = 0;
7      for (const ostos of ostokset) summa += ostos.hinta;
8      return summa;
9  }
    
```

Listaus 3.1: Esimerkki Closure-annotaatiosta funktiolle

TypeScript ja Flow puolestaan jatkavat ECMA-262 -spesifikaatiota erityisellä syntaksilla tyyppien eksplisiittistä määrittämistä varten.

```

1  function ostoskorinHinta(ostokset: Ostos[]): number {
    
```

Listaus 3.2: Esimerkki Flow tai TypeScript annotaatiosta funktiolle

Flow ja TypeScript -esimerkeissä tyyppiannotaatiot ovat osana koodia, mikä tekee ohjelmasta yhteensopimattoman tavallisen JavaScriptin kanssa. Ohjelma on käännettävä JavaScriptiksi ennen suorittamista, eikä annotaatioiden syntaksi ja merkitys ole suoraan selvä JavaScript-ohjelmoijalle.

Closuren annotaatiot on sijoitettu kommentteihin, joten niillä ei ole ajonaikaista vaikutusta ja ohjelma on täten sellaisenaan hyväksyttävää JavaScriptiä. Toisaalta kommentit voivat olla hankala ja runsassanainen formaatti monimutkaisille tyyppiannotaatioille, mikä kasvattaa niiden kirjoittamiseen vaadittua työmäärää [4, 8].

Aiemmassa esimerkissä esitelty tyyppi `Ostos` pitäisi määritellä Closurea varten muiden annotaatioiden tapaan dokumentaatiokommentteja käyttäen:

```
1 /**
2  * @typedef {}
3  *   hinta: number
4  * }}
5  */
6 let Ostos;
```

TypeScript ja Flow tarjoavat käännösaikaisen tyypin (type alias) määrittelyyn tiiviimmän ja helppolukuisemman syntaksin [4]:

```
1 type Ostos = {
2   hinta: number;
3 };
```

Tällainen tyypin määritteleminen vaikuttaa ainoastaan käännösvaiheen tyyppitarkastukseen, eikä määrittely tuota tietorakenteita tai muuta sisältöä suoritettavaan JavaScript-ohjelmaan.

3.2 Ohjelman kääntäminen ennen suorittamista

JavaScript koodi tulkitaan tai käännetään tyypillisesti suorittamisen yhteydessä, selaimesta tai muusta suoritussympäristöstä löytyvän ”moottorin” toimesta. EcmaScript standardin mukaista koodia suorittamaan suunnitellut moottorit, kuten V8 tai SpiderMonkey, eivät kuitenkaan osaa käsitellä TypeScript- tai Flow-annotaatioilla merkattua koodia. Näinollen TypeScript- tai Flow-annotoitu koodi on välttämätöntä kääntää muotoon jossa annotaatiot on poistettu ja jäljellä on enää standardinmukainen JavaScript. Koska JavaScriptin käyttäminen ei normaalisti vaadi erillistä käännösvaihetta, useissa projekteissa ei ole sellaista käytetty. Koodin minimointi ja muu optimointi on ollut parhaiden käytäntöjen mukaista jo tovin, mutta tällaiset koodinkäsittelyt tehdään yleensä vasta ennen ohjelman julkaisua. Kehittäjät ovat tavanomaisesti voineet suorittaa kirjoittamansa JavaScriptin sellaise-



Kuva 3.1: Vaihtoehtoisia käännösprosesseja

naan kehitysympäristössä. Käännösvaiheen aikavaatimus pyritään luonnollisesti pitämään mahdollisimman pienenä, mutta se on silti projektin monimutkaisuuteen ja kehitysnopeuteen vaikuttava tekijä joka työkalun käyttöönotossa tulee huomioida.

Kuvaaja 3.1 esittää erilaisia käännösprosesseja eri JavaScript-ohjelman kehityskielille. Jos koodi kirjoitetaan kaikkiin suoritussympäristöihin soveltuvalla kielen versiolla, kuten EcmaScriptin versiolla 5 (2011), käännösvaihe voidaan jättää kokonaan väliin. Kielen uudemmat versiot ovat kuitenkin tuoneet monia lisäominaisuuksia, joiden käyttäminen vanhempia selaimia tukevissa ohjelmissa vaatii joka tapauksessa käännösvaiheen staattisesta tyyppittämisestä riippumatta. Uusia JavaScript- ominaisuuksia hyödyntäville projekteille koodin kääntäminen ei siis tule täysin uutena vaiheena.

3.3 Työkalun vaiheittainen käyttöönotto

JavaScript-kirjastojen hallintaan tarkoitettu rekisteri npm on yksi suurimmista ohjelmistoekosysteemeistä [9]. Tutkinnon kirjoittamisen hetkellä rekisterin kotisivu, npmjs.com ilmoitti ladattavissa olevien pakettien määräksi yli 595 000. Avoimessa jakelussa olevien kirjastojen lisäksi JavaScriptiä käyttävillä kehittäjillä voi olla suuri määrä valmista JavaScript-koodia, jota voi hyödyntää uusissa projekteissa. Jotta TypeScript, Flow ja Closure olisivat hyödyllisiä työkaluja JavaScript ohjelmien kehitykseen, on niiden oltava yhteensopivia sellaisen JavaScript-koodin kanssa jonka kehitykseen ei kyseistä työkalua ole käytetty.

TypeScript ja Flow tukevat erityisiä määrittelytiedostoja, joiden sisältämällä tyyppianotoidulla koodilla määritetään kirjaston tai muun JavaScript koodin ulkoisen rajapinnan tyyppimäärittelyt. Näiden tiedostojen kirjoittamiseen käytetty syntaksi on sama kuin muissakin tiedostoissa, niiden funktio- ja metodimäärittelyistä on jätetty implementaatiot kokonaan pois. Tiedoston ei ole tarkoitus olla osana varsinaista suoritettavaa koodia, vaan se palvelee ainoastaan kuvauksena sellaisen koodin tyyppimäärittelystä, jonka käsittelyä TypeScript ja Flow eivät muuten voisi valvoa. TypeScriptillä voitaisiin esimerkiksi kirjoittaa seuraava tiedosto `ostoskori.d.ts`, jonka tehtävä on annotoida toista tiedostoa `ostoskori.js`.

```
1 /** Laittaa tuotteen ostoskoriin. */
2 export function lisääTuote(ostos: Ostos): void;
```

Listaus 3.3: Esimerkki TypeScript määrittelytiedostosta `ostoskori.d.ts`

Tämän jälkeen TypeScript tiedostosta käsin voidaan kutsua tätä JavaScript-funktiota siten, että TypeScript valvoo tyyppien oikeellisuutta.

```
1 import * as ostokset from './ostoskori';
2
3 ostokset.lisääTuote({ hinta: 5 });
```

Listaus 3.4: JavaScript koodin kutsuminen TypeScript tiedostosta `tuotesivu.ts`

Joissain tapauksissa kirjastoille tai kehittäjän omalle aiemmin kirjoitetulle JavaScript koodille ei kuitenkaan ole valmiita TypeScript-tyyppimäärittelyjä eikä niitä syystä tai toisesta voida luoda ennen muun kehityksen jatkamista. Kaikkien kolmen työkalun tärkeimpiin ominaisuuksiin kuuluu tuki vaiheittaiselle käyttöönotolle, eli käytännössä yhteenspiivuus täysin tyyppitarkastamattoman koodin kanssa. Sekä Flow että TypeScript tarjoavat erityisen yleisviittaustyyppin Any, jota voi käyttää kuvaamaan mitä tahansa JavaScript arvoa [4]. Any-tyyppiseen muuttujaan voidaan asettaa mikä tahansa arvo ja Any tyyppinen arvo voidaan asettaa mihin tahansa muuttujaan tai funktioparametriin. Any tyyppin avulla muuten staattisesti tyypitetystä ohjelmassa voidaan ohittaa käännösaikainen tyyppien tarkistaminen sellaisten koodin osien kohdalla joiden ajonaikaista arvoa olisi muuten vaikea tai mahdotonta määrittellä käännösaikana. Näin ollen, mikäli tarve vaatii täysin annotoimattoman JavaScript moduulin käyttämistä, kaikkien kyseisestä moduulista tuotujen arvojen voidaan määrittää olevan tyyppiä Any, jollin tarkastaja ei kiinnitä huomiota siihen miten JavaScriptillä määritettyjä funktioita tai muita arvoja käsitellään.

Luku 4

Virheiden havaitseminen

4.1 Virheiden havaitseminen

Kenties tärkein staattisen tyyppijärjestelmän tehtävä on havaita ja estää ohjelmoijan virheitä. Tässä esitelty työkalut, mahdollisesti Closure kääntäjää lukuunottamatta, onkin kehitetty erityisesti tätä tarkoitusta varten.

Kaikki kolme työkalua antaisivat käännösvirheen jos esimerkeissä 3.1 ja 3.2 esiteltyä funktiota kutsuttaisiin virheellisesti esimerkiksi listalla hintaa kuvaavia numeroita, sillä funktion parametrin on annotoitu olevan lista “Ostos”-tyyppimääritelmän mukaisia objekteja. Esimerkiksi virheellinen kutsu `ostoskorinHinta([5, 10, 15])` ei itse asiassa aiheuttaisi suoritettaessa ohjelman keskeyttävää virhettä. `ostos.hinta` ilmaisu on sallittu vaikka muuttuja `ostos` olisikin arvoltaan numero eikä objekti. Tällöin ilmaisun arvo on `undefined` ja lausekkeen `summa += ostos.hinta` jälkeen `summa` muuttujan arvo on erityinen ei-numeroa kuvaava `NaN` [10]. Käännösaikaisen tarkistamisen merkitys korostuu erityisen hyödylliseksi tämänkaltaisen ohjelmointivirheen kohdalla, sillä virhe ei välttämättä ole muutoin helposti havaittavissa. Funktiokutsu ei aiheuttaisi helposti todennettavaa suoritusajasta virhettä, joten ei-toivottu palautusarvo `NaN` saattaisi kiertää ohjelman operaatioiden välillä pitkällekin aiheuttaen muita loogisia virheitä.

Vuonna 2017 tehdyssä tutkimuksessa TypeScriptin ja Flown vaikutuksesta avoimen

lähdekoodin JavaScript-projekteihin havaittiin, että vähintään 15% ilmoitetuista ja korjatuista bugeista olisi voitu havaita ja välttää jos projektin kehitykseen oltaisiin käytetty jompaakumpaa näistä työkaluista [11]. Tutkimuksen arvioinnissa huomioitiin lisäksi, että tulos on tutkimusmenetelmästä johtuen mitä luultavimmin alempi kuin tällaisen muutoksen tuoma todellinen vaikutus.

4.2 Ohjelman optimointi käännösvaiheessa

TODO

4.3 Tyypimmääritys dokumentaationa

TODO

Luku 5

Ongelmat JavaScriptin staattisessa tyypittämisessä

5.1 Vaikeasti analysoitavat suunnittelumallit

5.2 Tyypitturvallisuus ja käytännöllisyys

Luku 6

Yhteenveto

Lähdeluettelo

- [1] Transition to OO programming - Safety and strong typing. URL <http://www.cs.cornell.edu/courses/cs1130/2012sp/1130selfpaced/module1/module1part4/strongtyping.html>.
- [2] JavaScript language resources. URL https://developer.mozilla.org/en-US/docs/Web/JavaScript/Language_Resources.
- [3] Standard ECMA-262 - ECMAScript 2017 Language Specification. URL <https://www.ecma-international.org/publications/standards/Ecma-262.htm>.
- [4] TypeScript Language Specification. URL <https://github.com/Microsoft/TypeScript/blob/b8fbf884d0f01c1a20bb921cc0a65d6c1a517ee8/doc/TypeScript%20Language%20Specification.pdf>, versio 1.8.
- [5] Installing and setting up Flow for a project. URL <https://flow.org/en/docs/install/>.
- [6] Closure Compiler. URL <https://developers.google.com/closure/compiler/>.

-
- [7] Annotating JavaScript for the Closure Compiler. <https://github.com/google/closure-compiler/wiki/Annotating-JavaScript-for-the-Closure-Compiler>.
- [8] Anders Hejlsberg. Microsoft Build 2014. TypeScript, maaliskuu 2014. URL <https://channel9.msdn.com/Events/Build/2014/3-576>.
- [9] Shriam Rajagopalan Erik Wittern, Philippe Suter. A Look at the Dynamics of the JavaScript Package Ecosystem. 2016.
- [10] Ecma International. *ECMA-262 Section 12.8.3*, kesäkuu 2017. URL <https://www.ecma-international.org/ecma-262/8.0/index.html#sec-additive-operators>, versio 8.0.
- [11] Earl T. Barr Zheng Gao, Christian Bird. To Type or Not to Type: Quantifying Detectable Bugs in JavaScript. 2017.